

Angela Borgwardt

Wissenschaftsbewertung – Wie kann sie reformiert werden?

 Eine Stunde für die Wissenschaft
Paper No. 10

AUF EINEN BLICK

Die Frage, wie Wissenschaft angemessener bewertet werden kann, wird gegenwärtig in Deutschland und im europäischen Kontext verstärkt diskutiert. Dahinter steht die Erkenntnis, dass zukunftsfähige Wissenschaft qualitätsbewusst und verantwortungsvoll sein muss, um ihre wichtigen Aufgaben für die Gesellschaft wahrnehmen zu können. Gegenwärtig zeigen sich bei der Wissenschaftsbewertung aber gravierende Defizite bedingt durch die Dominanz quantitativer und bibliometrischer Indikatoren, die der Vielfalt und den Inhalten wissenschaftlicher Leistungen nicht gerecht werden. Die damit verbundenen Fehlansätze wirken sich negativ auf das gesamte Wissenschaftssystem aus und gefährden die Qualität von Wissenschaft. Das vorliegende Papier skizziert die problematischen Auswirkungen des derzeitigen Systems, versammelt Vorschläge für eine adäquatere Wissenschaftsbewertung und gibt Empfehlungen für die Umsetzung von Reformen, die dazu beitragen können, dass wissenschaftliche Erkenntnisse der Gesellschaft besser als aktuell zugute kommen.

DAS SYSTEM DER WISSENSCHAFTSBEWERTUNG

Wissenschaftsbewertung beschäftigt sich mit der Frage, wie die wissenschaftlichen Leistungen von Individuen und Institutionen zu beurteilen sind. Die zugrunde gelegten Kriterien sind ausschlaggebend für die Zuschreibung von Reputation und die Verteilung von Ressourcen im Wissenschaftssystem. Damit werden auch die Inhalte und die Qualität der Wissenschaft beeinflusst, da sich die beteilig-

ten Akteur_innen an diesen Kriterien orientieren, um im System erfolgreich zu sein. Einen zentralen Stellenwert bei der Wissenschaftsbewertung hat die Forschungsbewertung, weil in vielen Bereichen des Wissenschaftsbetriebs die erbrachten Forschungsleistungen entscheidend sind, etwa bei den Karrierechancen an Universitäten oder bei der Vergabe von Fördermitteln.¹

Das gegenwärtige System der Wissenschaftsbewertung ist nur wenige Jahrzehnte alt und im Wesentlichen aus zwei Entwicklungen hervorgegangen (vgl. Dirnagl 2021): Zum einen war es eine Reaktion auf die massive Ausweitung wissenschaftlicher Forschung, die zu einer wachsenden Menge an Forscher_innen, Projekten, Anträgen und Publikationen geführt hat. Um den Bewertungsaufwand nicht zu groß werden zu lassen, wurden einfache und schnell zu erhebende Bewertungskriterien gebraucht. Zum anderen war der Wunsch nach Verteilungsgerechtigkeit und wissenschaftlicher Unabhängigkeit entscheidend: Wissenschaftsbewertung sollte auf objektiven Kriterien basieren, die nachvollziehbar und nicht willkürlich sind – und eine eindeutige Bewertung erlauben.

Diese Anforderungen schienen quantitative Indikatoren zu erfüllen, weil sie Leistungen messbar machen und relativ leicht einsetzbar sind. Sie sollten in erster Linie der Komplexitätsreduktion dienen und durch objektive Kriterien Vergleichbarkeit in einem Wissenschaftssystem ermöglichen, das von einem Wettbewerb um Ressourcen und Reputation gekennzeichnet ist.

¹ Die im Folgenden dargestellten Analysen und Empfehlungen basieren auf neueren wissenschaftlichen Studien und Stellungnahmen von Wissenschaftseinrichtungen zum Thema sowie auf den Inputs und der Diskussion im Rahmen einer Onlineveranstaltung der Friedrich-Ebert-Stiftung mit dem Titel „Was macht die Qualität von Wissenschaft aus? Möglichkeiten zur Reform der Wissenschaftsbewertung“, die am 23.3.2023 stattfand: <https://www.fes.de/themenportal-bildung-arbeit-digitalisierung/wissenschaft/eine-stunde-fuer-die-wissenschaft/was-macht-die-qualitaet-von-wissenschaft-aus> (13.6.2023).

Wachsende Bedeutung quantitativer Indikatoren

Bei der Bewertung von Wissenschaft werden inzwischen verschiedene quantitative Indikatoren in unterschiedlichen Verfahren eingesetzt. In der Forschung ist vor allem die Anzahl an Publikationen in bestimmten Fachjournals entscheidend. Eine besondere Rolle spielen dabei bibliometrische Indikatoren. Insbesondere die Bedeutung des Journal Impact Factor (JIF) ist in den vergangenen 20 Jahren stark gestiegen. Weitere kamen hinzu, zunächst der Hirsch-Faktor (h-Index) und in den letzten Jahren Altmetriken, die in ihrer Aussagekraft noch diskutiert werden (vgl. Franzen 2017; Kassab et al. 2020). Bibliometrische Indikatoren sind zu verbreiteten Bewertungsinstrumenten geworden, um die Forschungsleistungen von Wissenschaftler_innen und Institutionen, die Qualität und Relevanz von Fachpublikationen und Forschungsprojekten zu messen.

BIBLIOMETRISCHE INDIKATOREN

Bibliometrische Indikatoren basieren auf Maßzahlen, die entweder den Publikationsoutput (z. B. Anzahl von Publikationen) oder die Wahrnehmung in der Wissenschaftsgemeinde bzw. die Relevanz in der Öffentlichkeit messen (vgl. TU Dresden o. J.).

Der **Journal Impact Factor (JIF)** wurde in den 1960er Jahren entwickelt, um Bibliotheken eine Hilfestellung zu geben, welche Journals die größte Beachtung in der Fachcommunity finden und deshalb geführt werden sollten. Der JIF einer Zeitschrift gibt auf Basis einer Formel an, wie häufig die Artikel einer Zeitschrift in den vergangenen zwei Jahren in anderen Fachzeitschriften im Durchschnitt zitiert wurden. Die errechnete Zahl dient dem Vergleich verschiedener Zeitschriften und wird als Maß für die Reputation einer Zeitschrift interpretiert, indem ihr Einflussfaktor auf den wissenschaftlichen Diskurs bestimmt wird. Berechnungsgrundlage sind Zeitschriften, die in der multidisziplinären Zitationsdatenbank „Web of Science“ in der „Core Collection“ erfasst wurden.

Der **Hirsch-Index (h-Index)** wurde 2005 eingeführt und hat seitdem an Popularität gewonnen. Er ist eine Messgröße, um den Output von Forscher_innen und die weltweite Wahrnehmung ihrer Veröffentlichungen in Fachkreisen zu errechnen. Die Kennzahl beruht auf bibliometrischen Analysen von Zitationen der Publikationen eines Wissenschaftlers bzw. einer Wissenschaftlerin.

Altmetriken sind Kennzahlen, die ein breites Spektrum an Reaktionen im Internet auf eine wissenschaftliche Veröffentlichung quantifizieren und damit auch die Wirkung außerhalb der Wissenschaftsgemeinde erfassen sollen. Gemessen werden u. a. Aktionen, die ein Dokument im Internet aufrufen, herunterladen, besprechen und verlinken, auf Webseiten, Blogs und in sozialen Netzwerken diskutieren und liken.

Im deutschen Wissenschaftssystem wurde der Siegeszug der quantitativen Indikatoren bei der Bewertung wissenschaftlicher Leistungen auch dadurch unterstützt, dass in den 1990er Jahren Praktiken des unternehmerischen Handelns in das Hochschulwesen eingeführt wurden (vgl. Maasen/Weingart 1996: 19f.). Das Konzept der „unternehmerischen Universität“ orientierte sich an den Prinzipien Rechenschaftslegung, Transparenz und Effizienz und kann damit als Zäsur im Rollenverständnis der Hochschule gelten: Es markierte das Ende eines nicht systematisch hinterfragten Vertrauens in die Selbstregulierungsmechanismen der Wissenschaft bezüglich ihres sorgfältigen Umgangs mit öffentlichen Geldern und ihrer Ausrichtung auf das öffentliche Interesse. Das „Unternehmerische“ der Hochschulen löste einen „Boom von Quantifizierung und Evaluierung“ aus, der mit problematischen Entwicklungen einherging, etwa dass von Zahlen auf den Inhalt und die Qualität geschlossen wurde (Maasen/Weingart 1996: 42).

Diese Entwicklung hatte gravierende Folgen für das gesamte Wissenschaftssystem, da quantitative Messgrößen nun in fast allen Bereichen direkt oder indirekt zum Maßstab geworden sind und dort Steuerungswirkungen entfalten. Das betrifft beispielsweise:

- die Bewertung von Artikeln für wissenschaftliche Fachzeitschriften vor der Veröffentlichung;
- die hochschulinterne leistungsorientierte Mittelvergabe (LOM);
- die Beurteilung von Projektanträgen zur Forschungsförderung;
- die Vergabe von Drittmitteln und Auszeichnungen;
- die Leistungsbeurteilung in Berufungsverfahren und bei Karriereschritten;
- die Evaluierung und Rankings von Hochschulen und Forschungsinstitutionen.

Hoher Stellenwert von Peer-Review

Quantitative Indikatoren erzeugen Informationen, die direkt zu Entscheidungen führen können oder eine Grundlage für Peer-Review-Prozesse liefern. Peer-Review spielt bei der Bewertung wissenschaftlicher Leistungen eine zentrale Rolle: Dabei begutachten unabhängige Wissenschaftler_innen (sogenannte Peers = Gleichgestellte) mit einer spezifischen Expertise die Leistungen von Wissenschaftler_innen, von Projekten und Institutionen. Diese wissenschaftsinternen Begutachtungsverfahren sollen fundierte Qualitätsurteile über Wissenschaft ermöglichen, ihrer Vielfalt und Komplexität gerecht werden wie auch wissenschaftspolitische Entscheidungen unterstützen und legitimieren (vgl. Wissenschaftsrat 2011: 16). In Informed-Peer-Review-Verfahren wird das Urteil der Peers durch ein Set an quantitativen Indikatoren unterstützt, um einerseits subjektive (Vor-)Urteile auf den Prüfstand zu stellen und andererseits die – nicht ausreichend aussagekräftigen – Ergebnisse der quantitativen Indikatoren durch Expert_innen einordnen zu lassen.

Ein besonders wichtiges Peer-Review-Verfahren ist das Berufungsverfahren, da hier darüber entschieden wird, welche Wissenschaftler_innen an Hochschulen leh-

ren und forschen. Quantitative Messgrößen und bibliometrische Indikatoren können dabei helfen, eine große Menge an Forschungsleistungen schnell vergleichend zu bewerten. So ist es zum Beispiel zu einer verbreiteten Praxis geworden, dass Gutachter_innen in Berufungsverfahren weder den akademischen Lebenslauf² der Kandidat_innen noch einzelne, im Publikationsverzeichnis angegebene Artikel lesen, sondern nur einen Blick auf die Liste der Veröffentlichungen (mit JIF hinter den Journal-Namen) und die Höhe der eingeworbenen Drittmittel werfen. Auf dieser Basis kann sehr schnell eine Entscheidung getroffen werden.

PROBLEMATISCHE AUSWIRKUNGEN DES BEWERTUNGSSYSTEMS

Das Publizieren von Forschungsergebnissen hat bei der Wissenschaftsbewertung einen zentralen Stellenwert, wofür sich die Redewendung „Publish or Perish“ („Veröffentliche oder gehe unter“) durchgesetzt hat. Im Publikationswesen ist der JIF inzwischen die maßgebliche Messgröße in vielen Bewertungs- und Begutachtungsprozessen. Um in diesem System als qualitativ anerkannt zu werden, müssen Wissenschaftler_innen möglichst häufig in Zeitschriften mit einem hohen JIF veröffentlichen oder dort zitiert werden.

Mängel des Journal Impact Factor

Diese große Bedeutung des JIF steht jedoch in Widerspruch dazu, dass in verschiedenen Studien belegt wurde, dass diese Maßzahl für diesen Zweck völlig ungeeignet ist (z. B. Wissenschaftsrat 2011; Marx/Bornmann 2012). Der entscheidende Mangel liegt darin, dass der JIF auf ein Ranking von Fachzeitschriften ausgerichtet ist und somit nicht aussagekräftig für den Inhalt oder die Qualität einer einzelnen Forschungsleistung oder eines publizierten Beitrags sein kann. Weitere Nachteile sind, dass seine mögliche Höhe (und Reputation) stark von den (Teil-)Disziplinen abhängig ist, in denen unterschiedlich viel in Journalen publiziert wird, und dass er von Fachzeitschriften durch redaktionelle Strategien manipuliert werden kann, zum Beispiel indem Autor_innen aufgefordert werden, möglichst viele Artikel aus dem eigenen Journal zu zitieren. Zudem begünstigt der JIF selbstverstärkende Prozesse nach dem Matthäus-Prinzip („Wer hat, dem wird gegeben“), da bereits häufig zitierte Publikationen tendenziell auch zukünftig häufiger zitiert werden.

Der h-Index, der die Zitationszahlen der Forschungsarbeiten im fachspezifischen Diskurs erfasst, ist zwar für den einzelnen Artikel aussagekräftiger als der JIF, gibt aber ebenfalls keine Auskunft über die Qualität oder den Inhalt

² Der akademische Lebenslauf ist von zentraler Bedeutung bei Bewerbungen auf Professuren und wissenschaftliche Stellen, aber auch bei der Vergabe von Drittmitteln und Preisen. Er enthält nicht nur die wichtigsten Stationen zu Ausbildung und beruflicher Erfahrung, sondern auch eine Liste der Publikationen, der Lehr- und Vortragstätigkeiten sowie Angaben zu weiteren Tätigkeiten im wissenschaftlichen Kontext, z. B. Drittmittelinwerbungen, Mitgliedschaften in Gremien, Stipendien und Auszeichnungen, durchgeführte Forschungsprojekte.

einer Forschungsleistung. Hinzu kommt, dass er einer genauen Kenntnis des fachlichen Kontextes bedarf und die Beschaffung von Zitationsdaten zu einzelnen Publikationen meist aufwendig ist, was für eine große Menge an Forschungsartikeln kaum umsetzbar erscheint.

Folgen der Fehlanreize

Mit der starken Orientierung an quantitativen Indikatoren bei der Wissenschaftsbewertung sind Fehlanreize verbunden, die eine Zielverschiebung nach sich ziehen können: Wissenschaftler_innen streben dann nicht mehr eine möglichst gute Forschungsleistung an, sondern eine quantitative Steigerung im Hinblick auf die maßgeblichen Indikatoren (Wissenschaftsrat 2011: 20f.).

Die negativen Folgen zeigen sich besonders deutlich am Beispiel der impliziten Anforderung, möglichst viel in Zeitschriften mit hohem JIF zu publizieren. So werden Publikationsstrategien befördert, bei denen Wissenschaftler_innen nicht mehr primär das Ziel verfolgen, Forschungsergebnisse fachspezifisch sinnvoll zu vermitteln, sondern stattdessen versuchen, diese nach bestimmten Kriterien der Leistungsbewertung zu optimieren (vgl. z. B. Kaden 2022, Hagenström 2022). Dadurch entsteht – in den Disziplinen unterschiedlich stark ausgeprägt – ein hoher Publikationsdruck mit negativen Folgen. Das kann die bewusste Verzögerung der Veröffentlichung von Forschungsergebnissen in der Hoffnung auf eine (spätere) Publikation in einer Top-Zeitschrift sein, die Aufspaltung von Forschungsergebnissen in kleine Publikationseinheiten oder eine starke Tendenz zur Mehrfachverwertung eigener Texte („Selbstplagiate“ und Textrecycling). Letztlich gewinnen dann jene, die das Impact-Factor-Spiel am besten beherrschen (Fey/Osterloh 2017).

In diesem System bleiben zahlreiche Publikationsorgane und -formen ausgeschlossen, die für die Veröffentlichung von Forschungsergebnissen ebenfalls relevant sind, zum Beispiel internationale Journale außerhalb des „Web of Science“ oder Online-Veröffentlichungen. Zudem werden facherspezifische Unterschiede in der Publikationskultur nicht angemessen erfasst. Auch wenn in den meisten Disziplinen Forschungsergebnisse in Fachzeitschriften publiziert werden – in drei Viertel aller Fächer ist ihr Anteil hoch oder sehr hoch (DFG 2022: 12) –, sind doch erhebliche Unterschiede festzustellen: So werden beispielsweise in den Naturwissenschaften vor allem Artikel in Fachzeitschriften veröffentlicht, während in den Geistes- und Sozialwissenschaften Monografien und Sammelbände eine wichtige Rolle spielen.

Auch die Bewertungskriterien in Berufungsverfahren an Universitäten stehen in der Kritik: Durch eine Engführung der Leistungsanforderungen auf Forschung und die starke Orientierung an quantitativen Kriterien kann nicht mehr gewährleistet werden, dass die qualifiziertesten Bewerber_innen mit dem höchsten Potenzial berufen werden, die zugleich das breite Spektrum an wissenschaftlichen Leistungen abdecken können. Das birgt die Gefahr, dass zentrale Aufgaben wie Lehre und andere wissenschaftliche Tätigkeiten (z. B. Transfer, Wissenschaftskommunikation, Nachwuchsförderung) vernachlässigt werden, weil sie weniger

reputationsträchtigt oder karriereentscheidend sind. Schwächen in Berufungsverfahren können stark negative Effekte auf das gesamte Wissenschaftssystem haben, weil durch die Auswahl des Personals innerhalb einer Einrichtung und in der jeweiligen Fachgemeinschaft für lange Zeit Weichen gestellt werden.

Die starke Tendenz zur Quantifizierung und die Dominanz von Kennzahlen führt dazu, dass in Bewertungsprozessen die Inhalte, die Relevanz und die Qualität der Wissenschaft nicht mehr angemessen berücksichtigt werden (Dirnagl 2021). Das hat gravierende negative Auswirkungen im gesamten Wissenschaftssystem: Bei der Verteilung von Ressourcen und Reputation führen ungeeignete Kriterien zu falschen Entscheidungen und die öffentlichen Mittel für die Wissenschaft werden nicht adäquat oder ineffektiv eingesetzt (Dirnagl 2021). Die Dominanz von Quantität zieht somit einen Qualitätsverlust in der Wissenschaft nach sich und kann auch wissenschaftliches Fehlverhalten begünstigen (vgl. Borgwardt 2014).

Aus diesen Fehlanreizen können langfristige strukturelle negative Folgen im Wissenschaftssystem entstehen, etwa indem wichtige Forschungsfragen nicht bearbeitet werden, ungeeignete Bewerber_innen Stellen erhalten, qualifizierte Forscher_innen nicht gefördert werden und die Ausrichtung von Wissenschaftsinstitutionen von einseitigen Kriterien bestimmt wird. Um die vielfältigen Potenziale von Wissenschaft für die Gesellschaft auszuschöpfen, sind Verbesserungen in der Wissenschaftsbewertung deshalb unerlässlich.

VORSCHLÄGE FÜR EINE ANGEMESSENERE WISSENSCHAFTSBEWERTUNG

In den vergangenen zehn Jahren haben sich verschiedene Reforminitiativen gebildet, die von Wissenschaftler_innen und Akteur_innen der Wissenschaftsbewertung vorangetrieben werden und in öffentlichen Erklärungen und Vereinbarungen ihren Ausdruck gefunden haben. Dazu gehören z. B. die San Francisco Declaration on Research Assessment (DORA 2012), das Leiden Manifesto for Research Metrics (Hicks/Wouters et al. 2015), die Hong Kong Principles for Assessing Researchers (Moher/Bouter et al. 2020) und der Paris Call on Research Assessment (2022).

Auch im deutschen Wissenschaftssystem setzt sich zunehmend die Erkenntnis durch, dass das System der Forschungsbewertung reformiert werden muss (DFG 2022; Allianz der Wissenschaftsorganisationen 2022). Der Wissenschaftsrat hatte bereits 2011 eine Anpassung der bisherigen Bewertungs- und Begutachtungsverfahren angemahnt, um „gute Forschung zu ermöglichen und zu stärken“ (Wissenschaftsrat 2011). Ein Zuwachs an Quantität dürfe nicht pauschal mit einem Zuwachs an Leistung oder gar an Qualität gleichgesetzt werden. Um Fehlentwicklungen zu vermeiden, müsse die Steuerung von Forschung auf ein „breitgefächertes Repertoire an Instrumenten“ setzen und nicht auf „monetäre Anreize, die wenige Kennzahlen zugrunde legen“ (Wissenschaftsrat 2011).

Aus den Vorschlägen und Forderungen der vielfältigen Reforminitiativen können wesentliche Anregungen für

eine verbesserte Wissenschaftsbewertung herauskristallisiert werden.

Weitgehende Einigkeit besteht in der Auffassung, dass die bibliometrische Kennzahl JIF bei der Leistungsbewertung von Personen, Projekten und Institutionen nicht mehr eingesetzt werden sollte. Vielmehr sollte künftig die inhaltliche Qualität im Mittelpunkt von Bewertungs- und Begutachtungsprozessen stehen. Geeignete quantitative Indikatoren sollten ergänzend herangezogen werden.

Als weitere wichtige Aspekte werden benannt:

- Eine stärker inhaltliche Begutachtung wertet Peer-Review-Prozesse weiter auf: Es ist unerlässlich, dass Gutachter_innen mit Fachexpertise die Forschungsarbeiten der Wissenschaftler_innen lesen und diese in ihrer Qualität, Originalität und Relevanz beurteilen.
- Bewertungsprozesse sollten multidimensional angelegt sein, indem mehrere qualitative und quantitative Indikatoren zur Anwendung kommen, um die verschiedenen Aspekte wissenschaftlicher Qualität erfassen zu können.
- Da inhaltliche Qualität und Relevanz keine disziplinübergreifenden, statischen Größen sind, bedarf es fächerspezifischer Maßstäbe und Kriterien. Das betrifft sowohl qualitative wie auch quantitative Kriterien, die die Signifikanz und Bedeutung von Wissenschaft und Forschung adäquater als bisher abbilden sollten. Da sich die Kriterien je nach Bewertungsgebiet und Disziplin stark unterscheiden, sollten die jeweiligen Fachcommunitys in Zusammenarbeit mit Förderorganisationen und Wissenschaftsinstitutionen diese Kriterien gemeinsam entwickeln und definieren.
- Bei Publikationen sollte ein größeres Set an Formen und Orten der Veröffentlichung einbezogen werden, zum Beispiel Open-Access-Veröffentlichungen. Dabei sollten auch die unterschiedlichen Veröffentlichungspraktiken der Disziplinen berücksichtigt werden.
- Grundlage einer angemesseneren Bewertung ist eine genaue und transparente Berichterstattung über alle Phasen und Instrumente des Forschungsprozesses – von der Konzeption des Forschungsprojekts bis zur Bereitstellung der Ergebnisse, einschließlich des Forschungsdesigns, der angewendeten Methoden, der Datenerfassung, der Analyseverfahren und der Durchführungsschritte. Die Offenheit der Wissenschaft (offene Methoden und offene Daten, Open-Access-Veröffentlichungen) bildet die Grundlage für die Reproduzierbarkeit und Nachvollziehbarkeit von Forschungsergebnissen. Sie ist ein wichtiger Pfeiler für wissenschaftliche Integrität und Qualitätssicherung.
- Bei der Besetzung von wissenschaftlichen Stellen und Professuren sollten die Leistungen von Wissenschaftler_innen auf einer qualitativen Beurteilung der bisherigen Publikationen und des Lebenslaufs basieren. Neben Forschungsexzellenz sind weitere forschungsbezogene Leistungen zu würdigen, zum Beispiel die Erstellung von Forschungssoftware, Datenpflege oder das Engagement für Forschungsinfrastruktur. Je nach Stellenprofil muss die Vielfalt wissenschaftlicher Tätigkeiten in unterschiedlicher Gewichtung in die Bewertung ein-

gehen, etwa Leistungen in Lehre, Transfer, Mentoring, Begutachtung, Gremienarbeit, Wissenschaftskommunikation.

- Bei stellen- und förderpolitischen Entscheidungen sollten – abhängig von der Bewertungsaufgabe – weitere Aspekte einbezogen werden, um Wissenschaft zukunftsfähig zu machen. Dazu gehört Diversität in verschiedenen Dimensionen (vor allem Geschlechtergleichstellung, Internationalität), da Perspektivenvielfalt eine wichtige Voraussetzung für Qualität in der Wissenschaft ist. Von großer Bedeutung ist auch eine höhere Wertschätzung kollaborativer und kooperativer Praktiken sowie inter- und transdisziplinärer Herangehensweisen, da diese Ansätze bei der Bearbeitung der gesellschaftlichen Herausforderungen immer wichtiger werden.
- Das Peer-Review-Verfahren muss weiterentwickelt werden, um damit verbundenen Risiken entgegenzuwirken, etwa geschlechtsspezifischen und strukturkonservativen Verzerrungen (Bias). Auch die Dominanz bestimmter Denkschulen bei Auswahlverfahren kann dazu führen, dass nur „Gleichdenkende“ gefördert werden und ein Mainstreaming in der Forschungsförderung unterstützt wird (vgl. Neidhardt 2010). Es gilt, bei Peer-Review-Prozessen die Diversität von Gutachter_innengruppen (in Bezug auf Kompetenzen, Disziplinen, Theorien und Methoden, Karrierestadium, Geschlecht, Nationalität etc.) sicherzustellen und das Spektrum der Auswahlverfahren zu erweitern, etwa durch Zufallsauswahl oder ein Sondervotum (*wild card*) für Mitglieder von Begutachtungsgremien (vgl. Wissenschaftsrat 2017).

UMSETZUNG VON REFORMEN IN EUROPA

Bei der Reform der Wissenschaftsbewertung spielt auf europäischer Ebene das Agreement on Reforming Research Assessment (ARRA) (CoARA 2022b) eine zentrale Rolle. 2020 hatte die Europäische Kommission festgestellt, dass das derzeitige System der Forschungsbewertung auf ungeeignete Indikatoren ausgerichtet ist und den Inhalt wie auch den Mehrwert einer Veröffentlichung nicht ausreichend einbezieht (Europäische Kommission 2020). Im Anschluss beteiligten sich zahlreiche europäische Organisationen aus dem Bereich Wissenschaft und Forschung an der Erarbeitung einer Vereinbarung zur Reform der Forschungsbewertung, um die Qualität und Wirkung von Forschung zu erhöhen. In der 2022 veröffentlichten finalen Vereinbarung werden Ziele (Commitments) für eine Reform der Forschungsbewertung formuliert, in denen es im Kern um die Stärkung qualitativer Ansätze in der Forschungsbewertung und um die Anerkennung einer größeren Vielfalt wissenschaftlicher Praktiken und Beitragstypen geht. Am 1.6.2023 hatten 566 Wissenschaftsorganisationen aus aller Welt das Abkommen unterzeichnet.

AGREEMENT ON REFORMING RESEARCH ASSESSMENT (ARRA) 2022 – VEREINBARUNG ÜBER EINE REFORM DER FORSCHUNGSBEWERTUNG

Die Unterzeichnenden der Vereinbarung verpflichten sich zu einer gemeinsamen Vision mit vier Kernverpflichtungen (Core Commitments). Demnach soll Forschungsbewertung:

1. die Vielfalt der Beiträge und Karrieren in der Forschung anerkennen – in Übereinstimmung mit den jeweiligen Anforderungen und der Art der Forschung;
2. vorrangig auf einer qualitativen Beurteilung basieren, für die Peer-Review von zentraler Bedeutung ist, unterstützt durch einen verantwortungsvollen Gebrauch quantitativer Indikatoren;
3. auf die unangemessene Verwendung von journal- und publikationsbasierten Metriken verzichten, insbesondere auf den ungeeigneten Gebrauch von Journal Impact Factor (JIF) und h-Index;
4. Rankings von Forschungsorganisationen nicht unkritisch verwenden.

Darüber hinaus verpflichten sich die Mitgliedsorganisationen dazu:

- die notwendigen Ressourcen bereitzustellen, um den Kulturwandel in Gang zu bringen bzw. die Veränderungen in der Organisation umzusetzen;
- die Kriterien, Instrumente und Prozesse der Forschungsbewertung zu überprüfen und weiterzuentwickeln;
- das Bewusstsein für die Notwendigkeit einer Reform der Forschungsbewertung zu erhöhen und für eine transparente Kommunikation, Anleitung und Schulung in Bezug auf Bewertungskriterien und -prozesse sowie deren Verwendung zu sorgen;
- an einem Austausch über Praktiken und Erfahrungen teilzunehmen, um gegenseitiges Lernen innerhalb der Koalition und darüber hinaus zu ermöglichen;
- Fortschritte bei der Einhaltung der Prinzipien und deren Umsetzung zu kommunizieren;
- die Praktiken, Kriterien und Instrumente der Forschung auf der Basis von zuverlässigen Nachweisen und dem neuesten Stand der Forschung zu bewerten;
- die Forschungsdaten offen zugänglich zu machen, um das Sammeln von Daten und die Forschung nachvollziehbar zu machen (Open Science).

Aus dieser Initiative ging im Dezember 2022 die Coalition for Advancing Research Assessment (CoARA) hervor, ein Aktionsbündnis von Universitäten, Forschungsinstituten, Fördereinrichtungen, wissenschaftlichen Akademien und Fachgesellschaften. In einem strukturierten Prozess, organisiert durch das CoARA-Sekretariat, arbeiten die Mitgliedsorganisationen in Arbeitsgruppen in verschiedenen Themenbereichen daran, die Prinzipien der Wissenschaftsbewertung weiterzuentwickeln und in ihren Einrichtungen umzusetzen. ARRA soll ein Orientierungsgerüst in der Bewertungspraxis vorgeben, wenn in den Organisationen

neue Bewertungsverfahren etabliert werden, die an die jeweiligen Forschungsfelder, Disziplinen, Forschungsthemen und an den Zweck der Forschungsbewertung angepasst sind. So sollen Reformschritte hin zu einem Kulturwandel der Wissenschaftsbewertung verwirklicht werden. Ein konkreter Zeitrahmen gibt vor, dass die Mitgliedsorganisationen ein Jahr nach Unterzeichnung einen ersten Bericht über die eigenen Reformbemühungen gemäß eines Aktionsplans mit definierten „Milestones“ für die Einrichtung vorlegen. Nach fünf Jahren sollen die Mitgliedsorganisationen über die Ergebnisse des gesamten ersten Zyklus der eigenen Reformbemühungen Auskunft geben.



CoARA – Coalition for Advancing Research Assessment 2022

In der CoARA engagiert sich auch die Deutsche Forschungsgemeinschaft (DFG), die diese Initiative als einen wichtigen Schritt in die richtige Richtung betrachtet. Nach Auffassung von Prof. Dr. Matthias Koenig, Vizepräsident der DFG und Mitglied im Lenkungsausschuss der CoARA, harmonisiert die Zielsetzung der europäischen Initiative mit den bereits seit Langem etablierten Grundsätzen der DFG-Bewertungspraxis.³ Schon seit 2010 gelte etwa in Bezug auf Publikationsverzeichnisse das Prinzip „Qualität statt Quantität“, also eine Obergrenze bei der Anzahl der Publikationen, die im Anhang an DFG-Anträge oder im Lebenslauf angeführt werden. Koenig berichtet, dass die DFG die ersten beiden Kernverpflichtungen der Vereinbarung ARRA bereits verwirklicht und auch schon mit der Umsetzung der weiteren Verpflichtungen begonnen hat. Die DFG werde sich aktiv in die CoARA-Arbeitsgruppen einbringen, um den Prozess des Kulturwandels mitzugestalten. Leider sei in Deutschland das Engagement in der CoARA noch zurückhaltend, meinte Koenig. Einige wichtige Wissenschaftsakteur_innen hätten noch Vorbehalte – oftmals mit dem Argument, dass das Primat der wissenschaftlichen Exzellenz durch die Einführung weiterer qualitativer Kriterien nicht mehr hinreichend sichergestellt sei. Diese Auffassung basiere jedoch auf einer Fehlwahrnehmung der Zielsetzung der CoARA, bei der es gerade um eine bessere Messung der inhaltlichen Qualität von Forschung und eine angemessenere Abbildung der vielfältigen Beiträge zu hochqualitativer Forschung gehe. Die Initiative sei als partizipatives und inklusives Verfahren konzipiert, sodass ihre Reformmöglichkeiten stark davon abhängen, dass sich möglichst viele Mitgliedsorganisationen in den AGs engagieren.

In einem Positionspapier zum wissenschaftlichen Publizieren hatte die DFG 2022 konkrete Reformvorschläge für die Forschungsbewertung in die Debatte eingebracht (DFG 2022a). Die bisher dominierende Wissenschaftsbewertung anhand publikationsbasierter Metriken setze problematische Anreize, indem Quantität statt Qualität der Vorrang gegeben werde. Dadurch werde eine wissenschaftsadäquate Entwicklung des Publikationswesens und des Wissen-

schaftssystems insgesamt verhindert. Eine verantwortungsvolle Wissenschaftsbewertung, die die Qualitätssicherung von Wissenschaft einschließe, müsse sich auf ein breites Spektrum wissenschaftlicher Produktivität stützen und sich vorrangig an den Inhalten von Forschung orientieren.

Diesen Überlegungen ließ die DFG kurz darauf ein Maßnahmenpaket folgen (DFG 2022b). Die Bewertung der Leistung von Wissenschaftler_innen durch DFG-Gutachter_innen soll künftig in einer ganzheitlichen Betrachtung des individuellen wissenschaftlichen Werdegangs und auf Grundlage inhaltlich-qualitativer Kriterien erfolgen. Bei Forschungsförderanträgen wurde programmübergreifend eine obligatorische Lebenslaufvorlage eingeführt, die den Antragsteller_innen narrative und tabellarische Angaben ermöglicht, um besondere Lebensumstände oder zusätzliche Tätigkeiten in der Wissenschaft anzugeben, wie beispielsweise Grementätigkeiten oder der Aufbau einer Forschungsinfrastruktur. Auch das große Spektrum wissenschaftlicher Publikationsformen soll in Förderanträgen und Lebensläufen gewürdigt werden. Neben maximal zehn Publikationen in den häufigeren Publikationsformaten können nun bis zu zehn weitere Forschungsergebnisse in anderen Publikationsformaten aufgeführt werden, zum Beispiel Artikel auf Preprint-Servern oder Datensätze. Angaben zu quantitativen Metriken wie Impact-Faktoren und h-Indizes sollen bei der Begutachtung keine Berücksichtigung mehr finden. Mit diesen Modifizierungen und Neuerungen erhofft sich die DFG eine chancengerechtere, wissenschaftsadäquatere und qualitativ hochwertigere Begutachtungsgrundlage.

Herausforderungen

Mit der stärkeren Orientierung an Inhalten und qualitativen Kriterien in Bewertungs- und Begutachtungsprozessen sind neue Herausforderungen verbunden.

- Es muss sichergestellt werden, dass die große Menge an differenzierten Informationen noch verarbeitet und kompetent eingeschätzt werden kann. Es müssen handhabbare Mechanismen gefunden werden, die den Aufwand für die Gutachter_innen nicht zu groß werden lassen. Möglich wären zum Beispiel größere Begutachtungszyklen, aber auch eine kritische Debatte darüber, an welchen Stellen welche Art der Bewertung notwendig ist.
- Die Tätigkeit von Gutachter_innen wird zeitlich und inhaltlich herausfordernder. Schon heute ist es für Forschungsfördereinrichtungen und Fachjournale oftmals schwierig, geeignete Expert_innen in ausreichender Zahl zu finden, die zu einer ehrenamtlichen Begutachtungstätigkeit bereit sind. Deshalb muss auch nach Wegen gesucht werden, wie die Tätigkeit als Gutachter_in angemessener gewürdigt bzw. honoriert werden kann.
- In qualitativen Bewertungsprozessen muss der subjektiven Voreingenommenheit entgegengewirkt werden, um faire, transparente und wissenschaftsadäquate Entscheidungen zu erreichen. Dabei kann auf Erfahrungen und eingespielte Praktiken von Wissenschaftsinstitutionen und Forschungsfördereinrichtungen zurückgegriffen werden, auf das Sechs-Augen-Prinzip etwa oder mehrstufige Verfahren mit Begutachtungen und anschließenden

³ Impulsvortrag von Prof. Dr. Matthias Koenig, „Eine Stunde für die Wissenschaft“ am 23.3.2023, vgl. Fußnote 1.

den Paneldiskussionen, in denen eine Entscheidung kollektiv erarbeitet wird.

- Eine besondere Herausforderung besteht bei inter- und transdisziplinären Forschungsprojekten, da die Entscheidungen zwischen verschiedenen Scientific Communitys (mit häufig unterschiedlichen Kriterien) ausgehandelt werden müssen. Auch hier können bereits existierende Verfahren Orientierung bieten. So wurde zum Beispiel wissenschaftlich untersucht, wie in der Begutachtungspraxis interdisziplinärer Panels in den USA durch Kommunikationsprozesse Konsens über eine hohe wissenschaftliche Qualität und Originalität hergestellt und „akademische Exzellenz“ identifiziert werden kann (Lamont 2009).

EMPFEHLUNGEN FÜR DEN REFORMPROZESS

1 Orientierung der Bewertung an Inhalt und Vielfalt

Das Bewertungssystem in der Wissenschaft muss vor allem in zwei Aspekten verändert werden: Erstens muss eine Bewertungskultur etabliert werden, in der die inhaltliche Qualität von Wissenschaft die entscheidende Rolle spielt. Geeignete quantitative Indikatoren sollten unterstützend herangezogen werden. Die Qualität von Forschung kann nur adäquat beurteilt werden, wenn Fachgutachter_innen in Peer-Review-Prozessen die Inhalte, Methoden und Ergebnisse der Forschung konkret bewerten und den gesamten akademischen Lebenslauf der Wissenschaftler_innen einbeziehen. Zweitens muss mehr Raum geschaffen werden für eine größere Diversität des Outputs wissenschaftlichen Arbeitens, für unterschiedliche wissenschaftliche Karrieren und ein breiteres Spektrum der Leistungen von Wissenschaft in Bezug auf Personen, Projekte und Institutionen.

2 Aufklärung und gemeinsames Engagement für Reformen

Es gilt, in der wissenschaftlichen Community das Bewusstsein für den notwendigen Kulturwandel in der Bewertungs- und Begutachtungspraxis zu erhöhen und Vorbehalte unter Wissenschaftsakteur_innen abzubauen. Veränderungen im Bewertungssystem sollten von allen Akteur_innen, die an Wissenschaftsbewertung beteiligt sind, getragen und von diesen auch aktiv mitgestaltet werden. Die Coalition for Advancing Research Assessment (CoARA) auf europäischer Ebene bietet einen guten Rahmen, um den Kulturwandel voranzutreiben, indem sich Wissenschaftsakteur_innen austauschen, gemeinsam geeignete Kriterien und Verfahren entwickeln und in den einzelnen Wissenschaftsinstitutionen sukzessive umsetzen. Es ist wichtig, über die Ziele der CoARA aufzuklären und möglichst viele Organisationen und Institutionen zur Mitarbeit zu motivieren, weil eine angemessenere Kultur der Wissenschaftsbewertung nur auf Grundlage einer hohen Akzeptanz dieses Ansatzes in den wissenschaftlichen Communitys breit verankert werden kann.

3 Langfristiger Prozess und kontinuierliche Weiterentwicklung

Eine wirkliche Transformation der Wissenschaftsbewertung kann nur gelingen, wenn die neuen Impulse in Handlungsstrukturen und geeignete Verfahren übersetzt werden, die der Komplexität und den Zielen der Wissenschaftsbewertung angemessen sind. Das Agreement on Reforming Research Assessment (ARRA) kann hinsichtlich der Bewertungspraxis eine übergreifende Orientierung geben. Auf deren Grundlage können in den einzelnen Organisationen und Institutionen dann neue Bewertungsverfahren etabliert werden, die auf die jeweiligen Forschungsfelder, Disziplinen, Forschungsthemen und den Zweck der Forschungsbewertung zugeschnitten sind. Ein Kulturwandel der Bewertung im Wissenschaftssystem braucht viele institutionelle und strukturelle Veränderungen und ist als längerer Prozess zu denken, der schrittweise verwirklicht wird. Kriterien, Indikatoren und Verfahren der Wissenschaftsbewertung müssen stetig weiterentwickelt, das heißt, regelmäßig kritisch reflektiert und bei Bedarf modifiziert werden, um das Bewertungs- und Begutachtungssystem immer wieder an neue Entwicklungen in Wissenschaft und Gesellschaft anzupassen. ←

LITERATURVERZEICHNIS

Allianz der Wissenschaftsorganisationen 2022: Stellungnahme zur Initiative der Europäischen Kommission: Reform der Forschungsbewertung, https://www.wissenschaftsrat.de/download/2022/Allianz_Reform_Forschungsbewertung_2022_05_02.pdf?__blob=publicationFile&v=6 (5.5.2023).

Borgwardt, Angela 2014: Wissenschaft auf Abwegen? – Zum drohenden Qualitätsverlust in der Wissenschaft: Friedrich-Ebert-Stiftung, Berlin. <https://library.fes.de/pdf-files/studienfoerderung/11071.pdf> (26.4.2023).

CoARA (Coalition for Advancing Research Assessment) 2022a: Homepage, <https://coara.eu/> (1.5.2023).

CoARA 2022b: Agreement on Reforming Research Assessment, https://coara.eu/app/uploads/2022/09/2022_07_19_rra_agreement_final.pdf (1.5.2023).

DFG (Deutsche Forschungsgemeinschaft) 2022a: Wissenschaftliches Publizieren als Grundlage und Gestaltungsfeld der Wissenschaftsbewertung: Herausforderungen und Handlungsfelder, Positionspapier, Bonn, https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/publikationswesen/positionspapier_publikationswesen.pdf (4.5.2023).

DFG 2022b: Maßnahmenpaket zum Wandel der wissenschaftlichen Bewertungskultur, Information für die Wissenschaft Nr. 61, 1.9.2022, https://www.dfg.de/foerderung/info_wissenschaft/2022/info_wissenschaft_22_61/index.html (1.5.2023).

Dirnagl, Ulrich 2021: Back to the Future: Von industrieller zu inhaltlicher Forschungsbewertung, in: Laborjournal, 8.2.2021, https://www.laborjournal.de/rubric/narr/narr/m_n_21_01.php?consent=1 (3.5.2023).

DORA 2012: San Francisco Declaration on Research Assessment, publiziert auf der Jahreskonferenz der American Society for Cell Biology, <https://sf.dora.org/read/> (3.5.2023).

Europäische Kommission 2020: Mitteilung der Kommission an das Europäische Parlament, den Rat, den europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen: Ein neuer EFR für Forschung und Innovation, Brüssel, 30.9.2020, COM(2020) 628 final, <https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:52020DC0628&from=DE> (1.5.2023).

Franzen, Martina 2017: Digitale Resonanz: Neue Bewertungskulturen fordern die Wissenschaft heraus, Berlin, <https://bibliothek.wzb.eu/artikel/2017/f-20469.pdf> (10.4.2023).

Frey, Bruno S.; Osterloh, Margit 2017: Absurde Mess-Manie, in: Forschung & Lehre 10 (2017), <https://www.forschung-und-lehre.de/politik/absurde-mess-manie-136> (10.4.2023).

Hagenström, Felix 2022: Wiederverwertung eigener Texte in der Wissenschaft, <https://doi.org/10.5281/zenodo.7409616> (10.4.2023).

Hicks, Diana; Wouters, Paul et al. 2015: Bibliometrics: The Leiden Manifesto for Research Metrics, in: Nature 520, S. 429–431, <https://doi.org/10.1038/520429a> (10.4.2023).

Kaden, Ben 2022: Wissenschaftsbewertung und Open Access, 14.6.2022, <https://open-access-brandenburg.de/wissenschaftsbewertung-und-open-access-eine-einschatzung/> (10.5.2023).

Kassab, Omar; Bornmann, Lutz; Haunschild, Robin 2020: Can Altmetrics Reflect Societal Impact Considerations? Quantitative Science Studies 1 (2), S. 792–809, https://doi.org/10.1162/qss_a_00032 (20.4.2023).

Lamont, Michèle 2009: How Professors Think: Inside the Curious World of Academic Judgment, Cambridge, MA.

Maasen, Sabine; Weingart, Peter 2006: Unternehmerische Universität und neue Wissenschaftskultur, in: die hochschule 1 (2006), S. 19–45, https://www.hof.uni-halle.de/journal/texte/06_1/Maasen_Weingart_Unternehmerische_Universitaet.pdf (10.5.2023).

Marx, Werner; Bornmann, Lutz 2012: Der Journal Impact Factor: Aussagekraft, Grenzen und Alternativen in der Forschungsevaluation, Beiträge zur Hochschulforschung, 34. Jg., 2 (2012), https://www.wissenschaftsmanagement-online.de/sites/www.wissenschaftsmanagement-online.de/files/migrated_wimoarticle/2-2012-Marx-Bornmann.pdf (20.4.2023).

Moher, David; Bouter, Lex et al. 2020: The Hong Kong Principles for Assessing Researchers: Fostering Research Integrity, in: PLoS Biol 18 (7), <https://doi.org/10.1371/journal.pbio.3000737> (23.4.2023).

Neidhardt, Friedhelm 2010: Selbststeuerung der Wissenschaft: Peer Review, in: Simon, Dagmar; Knie, Andreas; Hornbostel, Stefan (Hrsg.): Handbuch Wissenschaftspolitik, Wiesbaden (2. Aufl. 2016), S. 280–292.

Paris Call on Research Assessment 2022: The Open Science Conference (OSEC) unter der Schirmherrschaft der französischen Präsidentschaft des Rates der Europäischen Union, <https://osec2022.eu/paris-call/> (20.4.2023).

TU (Technische Universität) Dresden o. J.: Ausgewählte bibliometrische Indikatoren, <https://tu-dresden.de/forschung-transfer/forschungsinformationen/bibliometrie/bibliometrische-indikatoren> (20.4.2023).

Wissenschaftsrat 2011: Empfehlungen zur Bewertung und Steuerung von Forschungsleistungen, Drs. 1656-11, Halle, 11.11.2011, https://www.wissenschaftsrat.de/download/archiv/1656-11.pdf?__blob=publicationFile&v=3 (23.4.2023).

Wissenschaftsrat 2017: Begutachtungen im Wissenschaftssystem, Positionspapier, Berlin, https://www.wissenschaftsrat.de/download/archiv/6680-17.pdf?__blob=publicationFile&v=4 (...)

AUTORIN

Dr. Angela Borgwardt ist Politikwissenschaftlerin und arbeitet als freie Publizistin und Moderatorin zu wissenschafts- und gesellschaftspolitischen Themen.

IMPRESSUM

ISBN: 978-3-98628-328-5

Juli 2023

© **Friedrich-Ebert-Stiftung**

Herausgeberin: Abteilung Analyse, Planung und Beratung
Godesberger Allee 149, 53175 Bonn

www.fes.de/apb

Für diese Publikation ist in der FES verantwortlich:
Florian Dähne

Bestellungen/Kontakt: apb-publikationen@fes.de

Satz: minus design, Berlin

Bilder Seite 1: picture alliance/Zoonar|Roman Budnikov, picture alliance/ZB|Jan-Peter Kasper, Collage: minus design

Die in dieser Publikation zum Ausdruck gebrachten Ansichten sind nicht notwendigerweise die der Friedrich-Ebert-Stiftung. Eine gewerbliche Nutzung der von der FES herausgegebenen Medien ist ohne schriftliche Zustimmung durch die FES nicht gestattet. Publikationen der Friedrich-Ebert-Stiftung dürfen nicht für Wahlkampfzwecke verwendet werden.