

Schmidt, Uwe; Deser, Annemarie

Nichtintendierte Effekte der Bewertung wissenschaftlicher Leistungen

Webler, Wolff-Dietrich [Hrsg.]: Überzogener und überhitzter Wettbewerb in der Wissenschaft I. Wissenschaftsförderung und ihre Irrwege. 2., überarbeitete Auflage. Bielefeld : UVW UniversitätsVerlagWebler 2024, S. 85-105. - (Hochschulwesen: Wissenschaft und Praxis; N. F. 43)



Quellenangabe/ Reference:

Schmidt, Uwe; Deser, Annemarie: Nichtintendierte Effekte der Bewertung wissenschaftlicher Leistungen
- In: Webler, Wolff-Dietrich [Hrsg.]: Überzogener und überhitzter Wettbewerb in der Wissenschaft I.
Wissenschaftsförderung und ihre Irrwege. 2., überarbeitete Auflage. Bielefeld : UVW
UniversitätsVerlagWebler 2024, S. 85-105 - URN: urn:nbn:de:0111-pedocs-308585 - DOI:
10.25656/01:30858; 10.53183/97839460175475

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-308585>

<https://doi.org/10.25656/01:30858>

Nutzungsbedingungen

Dieses Dokument steht unter folgender Creative Commons-Lizenz:
<http://creativecommons.org/licenses/by-sa/4.0/deed.de> - Sie dürfen das Werk bzw. den Inhalt vervielfältigen, verbreiten und öffentlich zugänglich machen sowie Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes anfertigen, solange sie den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen und die daraufhin neu entstandenen Werke bzw. Inhalte nur unter Verwendung von Lizenzbedingungen weitergeben, die mit denen dieses Lizenzvertrags identisch, vergleichbar oder kompatibel sind. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

This document is published under following Creative Commons-License:
<http://creativecommons.org/licenses/by-sa/4.0/deed.en> - You may copy, distribute and transmit, adapt or exhibit the work or its contents in public and alter, transform, or change this work as long as you attribute the work in the manner specified by the author or licensor. New resulting works or contents must be distributed pursuant to this license or an identical or comparable license.

By using this particular document, you accept the above-stated conditions of use.



Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

Nichtintendierte Effekte der Bewertung wissenschaftlicher Leistungen

Uwe Schmidt & Annemarie Deser

Quidquid agis semper agas et respice finem

1. Einführung

Die Bewertung von Leistungen ist in mehr oder weniger systematischer Form seit jeher genuin für Wissenschaft (Hüther/Krücken 2016). Die Bewertung von Publikationen und der damit verbundene kritische Diskurs neuer Erkenntnisse sind substanziell für wissenschaftlichen Fortschritt und für die Annäherung an Wahrheit durch Bestätigung oder Verwerfen von Konzepten und Ideen. Obgleich dies wissenschaftsimmanent ist, sind die dabei verwendeten Indikatoren, Maßstäbe und Verfahren umstritten. Hierzu trägt bei, dass die Bewertung wissenschaftlicher Leistungen nicht mehr nur im Kontext wissenschaftlicher Diskurse und Selektionsmechanismen selbst zum Tragen kommt, sondern auch zur Einschätzung und Herstellung von individueller und institutioneller Reputation dient sowie zur Verteilung von Mitteln und zur hochschulpolitischen Bewertung von Forscher:innen und Forschungseinrichtungen

Der vorliegende Beitrag befasst sich mit der Frage, inwieweit die Bewertung von wissenschaftlichen Leistungen neben intendierten Wirkungen auch nichtintendierte Effekte aufweist. Hierzu werden in einem ersten Schritt (Abschnitt 2) Überlegungen zum Phänomen nichtintendierter Effekte bzw. nichtintendierter Handlungsfolgen skizziert. Im Anschluss werden Verfahren und Kontexte vorgestellt, in denen die Bewertung wissenschaftlicher Leistungen von Relevanz ist. Im 3. Abschnitt werden einzelne Bewertungsindikatoren beschrieben und erörtert, um danach die Frage in den Blick zu nehmen, ob nichtintendierte Effekte in Abhängigkeit von den jeweiligen Bewertungsgrundlagen und Kontexten zu beobachten sind. Diese Frage wird sowohl grundsätzlich als auch exemplarisch am im Vereinten Königreich praktizierten und vieldiskutierten *Research Excellence Framework* (REF) erörtert.

2. Nichtintendierte Handlungsfolgen – Konzept und Implikationen

Das Konzept der nichtintendierten Folgen absichtsgeleiteten Handelns geht wesentlich auf die Arbeiten Robert K. Mertons in den 1930er Jahren zurück. Verbunden hiermit ist eine grundsätzliche Diskussion in der Soziologie zur Bedeutung des Phänomens nichtintendierter Effekte, bzw. dem Phänomen, dass geplante Handlungen in der Regel nie exakt die Effekte aufweisen, die beabsichtigt waren, was neben Merton u.a. von Popper als eine wesentliche Herausforderung der soziologischen Theoriebildung bezeichnet wurde (Wippler 1978, S. 155; Merton 1968, S. 120). Wie lassen sich nun nichtintendierte Effekte erklären und systematisieren?

Jedes Handeln hat Konsequenzen („the consequences result from the interplay of the action and the objective situation, the condition of the action” (Merton 1996, S. 176)). Sind die erwarteten Konsequenzen zugleich Motivation für das Handeln, sprechen wir von „absichtsgeleitetem Handeln“ (nur darum soll es an dieser Stelle gehen¹) und wir nennen diese Konsequenzen ‚Ziele‘ oder „intendierte Effekte“. Es liegt in der Natur der Dinge, dass selten eine unternommene Aktion nur einen einzigen Effekt (und dann auch noch den gewünschten) nach sich zieht; es kann auch in gänzlich anderen Bereichen als dem genuin adressierten zu Ergebnissen kommen – diese werden als „latente Effekte“ bezeichnet. In beiden Sphären nun (dem Zielbereich und dem der latenten Effekte) können die Konsequenzen positiv oder negativ im Sinne der handelnden Person sein (Abb. 1). Nichtintendierte Effekte sind mithin nicht zwingend damit gleichzusetzen, dass sie grundsätzlich als negativ oder im Nachhinein als unerwünscht wahrgenommen werden. Aus Perspektive des Handelnden sind zwar intendierte Effekte stets als zielführend zu interpretieren, doch ist denkbar, dass auch nichtintendierte Effekte als funktional oder positiv wahrgenommen werden können (Merton 1936, S. 895 und 1996, S. 176).

¹ Dies bedeutet, dass „er [Merton] die absichtsvolle Handlung, die Motive und somit eine Wahl zwischen zumindest zwei Alternativen des Handelns voraussetzt [und] vom bloßen Verhalten und vom habituellen Routinehandeln“ (Hönic 2020, S. 305, Anm. d. US & AD) abgrenzt.

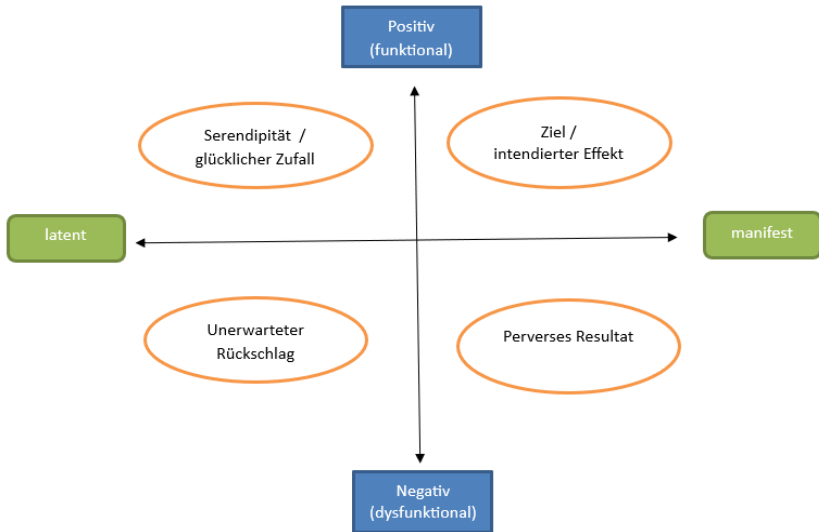


Abbildung 1: Effekte absichtsgeleiteten Handelns in Anlehnung an Merton 1968, S. 105, (eigene Darstellung)

Um die Einteilung an einem aktuellen politischen Beispiel zu verdeutlichen: Wenn eine Regierung im Falle einer Pandemie zum Schutze des Lebens der Bevölkerung Maßnahmen wie Lockdowns ergreift, wären die intendierten Effekte, dass aufgrund der geringen Ansteckung weniger Menschen als vorhergesagt der Krankheit erliegen², der Effekt allerdings, dass wegen Einschränkungen chronisch Kranke weniger gut medizinisch versorgt werden, kann als perverstes Resultat gewertet werden. Latent sind diejenigen Auswirkungen, die nicht direkt mit dem erklärten Ziel des Gesundheitsschutzes der Bevölkerungen zusammenhängen: so dass (als Serendipität) die Telearbeit zu- und die Umweltverschmutzung abnimmt³. Als negativ und latent könnten solche Folgen wie die Zunahme häuslicher Gewalt und die weiterreichenden Effekte auf die Volkswirtschaft gelten.⁴

Zu differenzieren ist darüber hinaus nach Merton zwischen *unorganized* and *formally organized actions*: „The first refers to actions of individuals considered distributive out of which may grow the second when like-minded individuals form an association in order to achieve a common purpose“ (Merton 1936, S. 896). Dementsprechend können Handlungsfolgen „nach den Folgen für Handelnde

² Es handelt sich also auch um eine „self-destroying prophecy“.

³ Womit also „undesired“ Effekte nicht zwingend „undesireable“ sind (Merton 1996, S. 176).

⁴ Merton erwähnt, dass es sich bei nichtintendierten Effekten keinesfalls um „nicht vorhergesehene“ Effekte handeln muss (obschon in der englischsprachigen Literatur „unintended“ und „un-anticipated“ synonym verwendet werden. Der intendierte Effekt einer Handlung kann mitunter, so Merton, auch das In-Kauf-nehmen des „kleineren Übels“ sein (Merton 1996, S. 176).

selbst oder für andere Personen, vermittelt durch die Sozialstruktur, die Kultur und die Zivilisation, unterschieden werden“ (Hönig 2020, S. 306).

Wie lassen sich nun nichtintendierte Effekte erklären und systematisieren? Nach Merton ist zu konstatieren, dass im Hinblick auf die Planung der Handlung die subjektive Definition der Situation entscheidend ist:

Men respond not only to the objective features of a situation, but also, and at time primarily, to the meaning this situation has for them. And once they have assigned some meaning of the situation, their consequent behavior and some of the consequences of that behavior are determined by the ascribed meaning” (Merton 1968, S. 175f.).

Dieser Effekt, als „Thomas-Theorem“ bekannt, bedeutet, dass sobald Menschen die Situation für sich als real und wichtig annehmen, auch die daraus folgenden Handlungen real werden (Merton 1968, S. 475). Weitere daran anschließende Kopplungen von Zuschreibung und Handlung sind der sogenannte Matthäus-Effekt⁵, d.h. dass denjenigen, die bereits spezifische Attribute erworben haben oder denen diese zugeschrieben werden, in der Folge auch mehr zugetraut und zugestanden wird, oder die selffulfilling-prophecy.

Merton unterscheidet fünf Gründe für das Auftreten nichtintendierter Effekte absichtsgeleiteten Handelns (vgl. auch Hönig 2020, S. 307):

(1) *Nichtwissen bzw. unvollständiges Wissen*

Nicht hinreichendes Wissen sei der häufigste Grund für die Erzeugung nicht-intendierter Effekte. Es führe dazu, dass Einschätzungen zu Konsequenzen des Handelns aus den Erfahrungen der Vergangenheit resümiert würden, was die Bewertung künftiger Handlungsfolgen nur unzureichend erfassen könne. Allerdings, so Merton, sei die Herstellung einer vollständigen Wissensbasis in der Regel auch aufgrund begrenzter Ressourcen nicht zu leisten im Sinne des „economic problem of distribution our fundamental resources, time and energy” (Merton 1936, S. 900, Hervorhebungen i.O.).

(2) *Fehleinschätzungen bzw. Irrtum*

Wenngleich an die Überlegungen zum Nichtwissen als Grund für nichtintendierte Effekte anknüpfend, sind hiervon Fehleinschätzungen zu unterscheiden. Diese beruhen auf dem Fehlschluss, dass bislang erfolgreiche Handlungsweisen generalisierbar und auf neue Situationen ohne Weiteres übertragbar sind: “But precisely because habit of a mode of activity which has previously led to the attainment of certain ends, it tends to become automatic and deliberative though continued repetition so that the actor fails to recognize that procedures which have been successful in certain circumstances need not be so under any and all conditions.” (Merton 1936, S. 901)

⁵ Diese systematische Beschreibung verdanken wir ebenfalls Robert Merton (Merton 1996, Kapitel 24 ab S. 318).

(3) *Unmittelbarkeit des Interesses*

Unter der Unmittelbarkeit des Interesses ist zu verstehen, dass spezifische Handlungsziele als Priorität definiert und damit andere, ggf. konkurrierende Aspekte ausgeschlossen werden – oder mit den Worten Mertons: „where the actor's paramount concern with the foreseen immediate consequences excludes the consideration of further or other consequences of the same act“ (Merton 1936, S. 901). Ein aktuelles Beispiel hierfür kann in der Fokussierung politischen Handelns während der Corona-Pandemie auf die Funktionsfähigkeit des Gesundheitssystems gesehen werden, womit parallele Herausforderungen bspw. im Bildungssystem als nachgeordnet behandelt wurden.

(4) *Grundlegende Werte*

Dieser von Merton benannte Grund für nichtintendierte Handlungseffekte rekurriert auf die Dominanz von Werthaltungen, die keine alternativen Handlungsoptionen bzw. eine Bewertung der Situation und der Handlungsziele außerhalb eines spezifischen Wertekanons zulassen. Beispielhaft verweist er auf die Ausführungen von Weber zur protestantischen Ethik und sieht darin ein grundlegendes Paradox, da die konsequente Handlungsorientierung an Werthaltungen nicht zwingend situationsadäquat ist und die zugrunde liegenden Werte selbst gefährdet: „Here is the essential paradox of social action – the ‚realization‘ of values may lead to their renunciation.“ (Merton 1936, S. 903)

(5) *„Selbst-zerstörende Prophezeiung“*

Hierunter versteht Merton den Einfluss von (öffentlichen) Vorhersagen, die individuelle Handlungspräferenzen ggf. auch entgegen rationaler Erwägungen beeinflussen: „Public predictions of future social developments are frequently not sustained precisely because the prediction has become a new element in the concrete situation, thus tending to change the initial course of developments.“ (Merton 1936, S. 903f.) Ein Beispiel hierfür wäre, wenn einem Schüler vermittelt wird, dass eine Prüfung kaum oder nicht erfolgreich von ihm zu absolvieren sei, was dazu führt, dass er seine Ambitionen, die Prüfung zu bestehen, aufgibt und sich nicht vorbereitet. Umgekehrt könnte er allerdings auch im Sinne einer self-fulfilling-prophecy durch die Vorhersage, dass er sicher die Prüfung bewältigen kann, ermuntert werden, sich intensiv vorzubereiten.

Diese kurzen zusammenfassenden Überlegungen zu Mertons Konzept und insbesondere den Gründen für nichtintendierte Handlungseffekte sollen – soweit möglich – als ein Rahmen bzw. eine Heuristik für die in den Abschnitten 5 und 6 dieses Beitrags ausgeführten Beispiele für nichtintendierte Effekte im Wissenschaftssystem dienen.

3. Verfahren und Kontexte der Bewertung wissenschaftlicher Leistungen

Die Bewertung wissenschaftlicher Leistungen erfolgt in unterschiedlichen Kontexten und mit divergenten Zielsetzungen. Grundsätzlich dient sie der Einschätzung wissenschaftlicher Leistungen innerhalb des Wissenschaftssystems und ist assoziiert mit der Idee der Selbststeuerungsfähigkeit desselben und damit auch der Verteilung von Partizipationschancen am wissenschaftlichen Diskurs. Dies findet seinen Ausdruck insbesondere in der Bewertung von Publikationen, von Projektanträgen oder Institutionen, bspw. in Form der Evaluation von Fächern, Fakultäten oder Fachbereichen. Zweck ist zum einen die Wahrung wissenschaftlicher Standards im Hinblick auf das jeweilige Selbstverständnis des Fachs. Zum anderen wird damit die Selektion der besten wissenschaftlichen Leistungen, so etwa in Form von Projektanträgen oder Publikationen verfolgt, was allerdings besonderen methodischen Herausforderungen unterliegt, worauf im Verlaufe des Beitrags noch eingegangen wird.

Die zweite Perspektive, die mit der Bewertung wissenschaftlicher Leistungen verbunden sein kann, ist jene der individuellen oder institutionellen Reputation. Zu nennen sind hier vor allem bibliometrische Daten, wie Anzahl der Publikationen und Zitationen, die Anzahl und der Verbleib des wissenschaftlichen Nachwuchses, Wissenschaftspreise oder, auf institutioneller Ebene, Rankings und Ratings. Diese Bewertungen wiederum kreieren Chancen – sei es im Zusammenhang von kompetitiven Verfahren wie der Exzellenzinitiative des Bundes und der Länder oder mit Blick auf wissenschaftliche Karrieren.

Der dritte Kontext schließlich, in dem die Bewertung wissenschaftlicher Leistungen von Relevanz ist, betrifft die Verteilung von Mitteln durch die zuständigen Ministerien bzw. die politische Steuerungsebene, wie auch die Verteilung der Mittel innerhalb der Hochschulen. Dabei dominiert der Bildungsauftrag der Hochschulen, das heißt, dass Mittel wesentlich entlang der Studierendenzahlen verteilt werden. Der finanziellen Goutierung von Forschungsleistungen kommt in diesem Zusammenhang eher ein gewisser „Bonus“-Charakter zu, wenngleich in einzelnen Bundesländern und innerhalb von Hochschulen auch Forschungsindikatoren – in der Regel Drittmiteleinwerbung und Anzahl der Promotionen sowie eingeworbene kollaborative Forschungscluster – zum Tragen kommen.

Bezieht man die vorherige grobe Einteilung der Zielsetzungen der Bewertung auf die jeweiligen Funktionen, die diese bzw. Evaluation wissenschaftlicher Leistungen einnehmen können, so kann man – unter Zugrundelegung des Verständnisses von Stockmann bzw. Stockmann/Meyer – nach Rechenschaftslegung, Entwicklungsfunktion, Kontrolle und Forschungsfunktion differenzieren (Stockmann 2007, S. 35f.; Stockmann/Meyer 2014, S. 82ff.). Hieraus ist zu schließen, dass die ‚intrawissenschaftlichen‘ Verfahren ein gewisses Maß an Kontrollfunktion, hauptsächlich aber eine Entwicklungsfunktion einnehmen. Dies bedeutet, dass es zum einen darum geht, die individuelle wie auch kollektive Leistungsfähigkeit auf Fachebene oder auch fachübergreifend zu erfassen und vergleichend zu be-

werten und andererseits Impulse für die fachliche und strukturelle Weiterentwicklung zu geben. Auch mit Blick auf die Herstellung individueller Reputation sind die Kontroll- und Entwicklungsfunktion latent, jedoch lassen sich diese Effekte besser durch die Erweiterung der Stockmann'schen Funktionen durch Schmidt & Lenz umschreiben, welche die Reputationsfunktion direkt adressieren und die genannten Funktionen durch jene der Mobilisierung, der Reputation und der Selbstvergewisserung ergänzen (Schmidt/Lenz 2021). Für die institutionelle Bewertung hat in diesem Sinne die Bewertung wissenschaftlicher Leistungen bspw. durch Rankings oder Evaluationen einen mobilisierenden Effekt. Die bei Schmidt & Lenz beschriebene Funktionendifferenzierung lässt sich auch auf die Mittelverteilung auf Grundlage der Bewertung wissenschaftlicher Leistungen beziehen. Hier steht neben der Kontrolle und Rechenschaftslegung häufig die Funktion der Selbstvergewisserung im Raum. Hierunter ist entsprechend den Überlegungen von Schwarz (2006) zu verstehen, dass mit der gewachsenen Autonomie von Bildungseinrichtungen die Notwendigkeit, intern gerechte oder als gerecht empfundene Steuerungsmechanismen zu entwickeln, einhergeht. Dies ist mit Blick auf die besonderen Herausforderungen von Universitäten als Expertenorganisationen (vgl. zusammenfassend Musselin 2007) nicht selbstredend und bedarf der organisationalen Selbstvergewisserung, um hinreichend Akzeptanz herzustellen, in der Regel durch Evaluationen bzw. andere Bewertungsverfahren.

Verfahren zur Bewertung wissenschaftlicher Leistungen lassen sich über die vorherigen Ausführungen hinaus grob nach solchen differenzieren, die datenbasiert sind, und solchen, die im Sinne eines Peer Reviews auf Einschätzungen von Fachkolleg:innen abstellen. Rein datenbasierte Verfahren spielen vor allem im Rahmen der Verteilung von Mitteln oder bei vergleichenden Bewertungen eine Rolle. In der Mehrzahl der Bewertungsverfahren jedoch – vor allem in solchen, die wissenschaftsimmanent zu verstehen sind – dominieren Peer Reviews. Mit Blick auf die strukturell-inhaltliche Steuerung des Wissenschaftssystems findet in der Regel ein Hybrid in Form eines sogenannten „Informed Peer Reviews“ Anwendung. Dies bedeutet, dass Peer Reviews auf einer hinreichenden Basis quantitativer und auch qualitativer Daten durchgeführt werden. Entsprechend der unterschiedlichen nationalen Kulturen und Steuerungssysteme lassen sich die Verfahren nach Sivertsen (2017, S. 78) im Hinblick auf die Ressourcenallokation in vier Kategorien einteilen:

- (1) die genannten Informed Peer Reviews, wie sie u.a. in England, Italien und Portugal Verwendung finden,
- (2) rein auf Forschungsindikatoren gestützte Systeme (u.a. Kroatien, Polen und Schweden),
- (3) solche, die neben Forschungsindikatoren noch weitere Leistungsbereiche, also insbesondere Lehrleistungen berücksichtigen (u.a. Dänemark, Norwegen und Finnland) und schließlich
- (4) jene, die ebenfalls ein breites Leitungsspektrum berücksichtigen, aber auf die Nutzung bibliometrischer Daten verzichten (u.a. Österreich und Niederlande)

Rankings, die im Wesentlichen auf eine quantitative Datengrundlage rekurrieren, sind das *Academic Ranking of World Universities* (Shanghai Ranking), das *Leiden Ranking*, das *QS World University Ranking* oder das *U Multirank*. Hierbei kommen Indikatorensets zur Anwendung, die unterschiedliche Aspekte fokussieren: Publikationsoutput und Impact (insbesondere durch Zitationen), aber auch Ausbildungsindikatoren, wie Betreuungsrelationen oder – wie das Shanghai Ranking – die Anzahl der Nobelpreisträger:innen als indirekter Qualitätsnachweis für eine gute Ausbildung.

4. Indikatoren zur Bewertung wissenschaftlicher Leistungen

In welchem Maße – wissenschaftlich fundierte – Daten zur Bewertung von Forschungsleistungen im Hinblick auf die skizzierten Verfahren Eingang finden und insbesondere welche Art von Leistungen damit abgebildet werden, stellt eine anhaltende Diskussion dar. Diese fokussiert vor allem auf Fragen der grundsätzlichen Angemessenheit der verwendeten Indikatoren, aber auch auf die Validität und Reliabilität der Daten. Folgt man den bereits Ende der 1990er Jahre durch Hornbostel (1997, S. 195ff.) dargelegten Einteilung von Indikatoren, so lassen sich sechs Indikatoren bzw. Indikatorengruppen identifizieren:

- Peer Reviews, also die Bewertung durch Fachkolleg:innen,
- Wissenschaftspreise,
- Drittmittelinwerbungen, die in den vergangenen Jahren enorm an Bedeutung gewonnen haben und an vielen Universitäten inzwischen mehr als ein Drittel des Gesamtetats ausmachen,
- Publikationsanalysen, also die Anzahl der Veröffentlichungen, ggf. differenziert nach Typen (Artikel nach Begutachtungsverfahren, Monographien, Herausgeberschaften usw.) und damit verbundener Gewichtung, die in der Regel fachspezifisch divergiert,
- Zitationsanalysen in unterschiedlichen Ausprägungen, d.h. die Anzahl der Zitationen, die durchschnittliche Zitationshäufigkeit pro Publikation, der sogenannte Hirsch-Index⁶, und
- Strukturindikatoren, wobei diese an sich nicht als Leistungsindikatoren bezeichnet werden können, sondern spezifische Muster, bspw. von Texten, auf einer deskriptiven Ebene erfassen.

In den vergangenen Jahren sind mit Blick auf Wissenschaftsindikatoren zwei Entwicklung zu beobachten: Zum einen werden insbesondere im Bereich der Publikations- und Zitationsanalysen weitere Differenzierungen vorgenommen, die

⁶ Also die Anzahl der Publikationen, die jeweils eine bestimmte Anzahl an Zitationen erreicht haben (ein Hirsch-Index von 10 bedeutet, dass mindestens 10 Publikationen zehnmal zitiert wurden)

bspw. auf Kennzahlen fokussieren, die den Anteil von Publikationen an einer Institution ausweisen, oder zu den 10% am häufigsten zitierten gehören. Eine zweite Entwicklung trägt der Beobachtung Rechnung, dass Publikationen zum Teil nicht mehr in klassischen Formaten, sondern als open science-Veröffentlichungen vorliegen und zudem nicht ausschließlich im wissenschaftlichen Diskurs rezipiert, sondern auch darüber hinaus wahrgenommen werden. Sogenannte Altmetrics (aus *alternative & metrics*) befassen sich entsprechend mit der Frage, wie wissenschaftlicher Output im gesellschaftlichen Umfeld wahrgenommen wird. Diese können im Vergleich zur Rezeption von Publikationen im Wissenschaftssystem den Vorteil haben, dass die ‚messbare‘ Beachtung und Diskussion zeitlich näher am Prozess wissenschaftlicher Produktion verortet sind.

5. Intendierte und nichtintendierte Effekte der Bewertung wissenschaftlicher Leistungen

Bewertungen haben in der Regel immer auch einen Einfluss auf das Verhalten der Akteure, das heißt, dass diese Bewertungen nicht nur ‚passiv‘ wahrnehmen, sondern aktiv und strategisch reagieren. Beobachtung und Bewertung sorgen also für eine Veränderung der beobachteten oder bewerteten Situation. In dem Dilemma des „Eingreifens“ in die zuvor „objektive“ Realität durch den Blick der sozialwissenschaftlichen Forschung sieht Merton Ähnlichkeiten mit dem Beobachtungsparadox in den Naturwissenschaften, wie es von Heisenberg beschrieben wurde (Merton 1936, S. 903-904; Merton 1968, S. 184). Bewertungen von wissenschaftlichen Leistungen stellen also nie eine quasi ‚neutrale‘ Beobachtung dar, sondern sind selbst ein wesentlicher Faktor für die Definition dessen, was als gut oder erfolgreich verstanden wird. Im Extremfall kann eine Bewertung oder Definition einer Situation diese derart verändern, dass es nach Merton zu Phänomenen der self-destroying- oder umgekehrt zur self-fulfilling-prophecy kommt.

Systemische Effekte durch die Bewertung wissenschaftlicher Leistungen sind auch durch spezifische Werthaltungen der Scientific Community oder der Hochschulpolitik zu erwarten, die bspw. eine hohe Präferenz für kollaborative, im besten Fall interdisziplinäre Forschungsvorhaben nahelegen und damit ggf. insofern dysfunktional wirken können, als sie individuelle Forschung als weniger wertvoll kategorisieren, obgleich sie in manchen Kontexten zielführender zu sein vermag. Indiz hierfür ist die hohe Bedeutung für kollaborative Forschungsvorhaben in Form von Exzellenzclustern, Sonderforschungsbereichen oder Graduiertenkollegs.

Gleichzeitig adaptieren Forschende zunehmend diese in der Regel als extern wahrgenommenen Erfolgskriterien und folgen den damit nahegelegten Handlungslogiken, wie bspw. spezifischen Publikationsstrategien. Häufig genannte Beispiele sind in diesem Zusammenhang die sogenannte ‚Salami-Taktik‘, d.h. die Aufspaltung von Forschungsergebnissen in möglichst viele Publikationen, die jeweils auf Teilaspekte eingehen und die Orientierung an Journals mit hohem

Impact Faktor oder Publikationen in Open Science, um einfacher auffindbar zu sein. Die damit verbundenen nichtintendierten Effekte können unterschiedlicher Art sein: Mit Blick auf die ‚Salami-Taktik‘ ist eine Partialisierung von Wissen zu erwarten, das ggf. den Blick auf weiterreichende Zusammenhänge der Forschungsergebnisse verstellt. Die Orientierung an Zeitschriften mit hohem Impact Faktor führte ungeachtet der grundsätzlichen Kritik an der Generierung und Verwendung des Journal Impact Faktor (Marx/Bornmann 2012) in den vergangenen Jahren in einigen Fachgebieten zu einer Zunahme sogenannter Fake-Verlage, die einen vergleichsweise hohen Impact Faktor ausweisen, gleichzeitig aber keine angemessenen qualitätssichernden Verfahren, d.h. insbesondere kein belastbares Peer Review nutzen. Bei der Publikation in Open Access-Formaten schließlich ist als nichtintendierter Effekt festzuhalten, dass die in der Regel hohe finanzielle Selbstbeteiligung der Autor:innen dazu führt, dass die Wahrnehmung von Forschungsergebnissen maßgeblich von den zur Verfügung stehenden finanziellen Ressourcen abhängen kann. Die durch Open Access angestrebte breite Verfügbarkeit von Forschungsergebnissen wird damit gewissermaßen käuflich und läuft den ursprünglichen Intentionen des Open Access-Gedankens zuwider. Diese Beispiele können in Mertons Verständnis aus Perspektive der Forschenden sowohl als fehlendes Wissen als auch als falsche Einschätzung über Wirkzusammenhänge ihrer Publikationsstrategie mit den damit verbundenen Effekten diagnostiziert werden. Sie haben insbesondere Effekte auf Ebene des Wissenschaftssystems, indem veränderte Anreize kreiert werden, die ihrerseits zu einer veränderten Definition der Situation und damit des wahrgenommenen Handlungsrahmens für Forschende führen.

Im Folgenden sollen am Beispiel des in Großbritannien verwendeten Research Excellence Framework (REF) nochmals Effekte auf Handlungs- und Systemebene skizziert werden, um die damit verbundenen komplexen Interdependenzen der unterschiedlichen Ebenen im Zusammenhang darzustellen.

6. Das Research Excellence Framework (REF)

6.1 Definition

Das REF ist ein UK-weites Bewertungsverfahren wissenschaftlicher Leistungen zwecks Mittelzuweisung an Higher Educational Institutions (HEI). Gemeinsam getragen von Fördermittelgebern in England, Schottland, Wales und Nordirland, bildet es zusammen mit den monetären Zuwendungen für Projekte („project-specific grants“) das „dual support system“ der staatlichen Mittelallokation für den HE-Sektor (Stern 2016, S. 6) und fokussiert auf die strategische und innovative Ausrichtung der HEI (strategic institutional funding).

Erstmals 1986 durchgeführt mit dem Ziel, die damals sehr knappen Ressourcen auf transparentem Wege in die exzellentesten Forschungsdepartments und Wissenschaftler:innen zu investieren, hat es sich mittlerweile etabliert: zwar unre-

gelmäßig (zwischen den bisher 8 Durchläufen liegen 3-8 Jahre jeweils), unter verschiedenen Namen (zunächst „Research Selectivity Exercise“, dann bis 2008 „Research Assessment Exercise“ genannt) und mit Anpassungen der Leistungsmessungsmethoden (Sivertsen 2017, S. 2; Hamann 2021, S. 84; Pinar/Horne 2022, S. 173), hat es einen festen Platz in der Wissenschaftslandschaft des Vereinten Königreichs. Es zählt zu einem der ältesten, komplexesten, aufwendigsten und stark institutionalisierten Forschungsassessments weltweit (und auch zu den am meist untersuchten (Marques et al. 2018)), und es hat auch eine große Strahlkraft auf Mittelzuweisungsverfahren anderer Länder entfaltet (Stern 2016, S. 6; Sivertsen 2017, S. 2). Es nimmt nicht wunder, dass die Kosten für die Durchführung beachtlich sind (Stern 2016, S. 6), wobei diese nach Einschätzung des UK Research and Innovation (UKRI)⁷ in Relation zu den verausgabten Mitteln moderat seien: Unmittelbar vergeben die vier UK-funding bodies aufgrund des REF-Rankings geschätzte zwei Milliarden £ jährlich an UK-Universitäten⁸; etwa 1% davon, so schätzt das REF-Team selbst, entfällt auf die Durchführung des REF (UK Research and Innovation (UKRI) 2022, S. 2). Das Ergebnis des letzten REF 2021 weist im weltweiten Vergleich eine hohe Qualität der HEI in UK aus: von den 1.878 teilnehmenden Instituten wurden 41% als „world-leading“ klassifiziert, weitere 43% als „internationally excellent“ (UK Research and Innovation (UKRI) 2022, S. 1).

Um am REF teilzunehmen, müssen Departments aus ihren Reihen Wissenschaftler:innen nominieren, diese wiederum reichen vier ihrer „research outputs“ ein, deren Veröffentlichung in den jeweiligen REF-Bewertungszeitraum fallen (Marginson 2015, S. 12; Hamann 2021, S. 84). Die Publikationen werden in den fachspezifischen Panels (Unit of Assessments) durch Peer-Review in Form von fünf Stufen (0-4 Sterne) bewertet und als zusammengefasste Note auf Ebene des Departements ausgewiesen (Hamann 2021, S. 85). Seit 2014 werden nicht nur Publikationen in der Benotung berücksichtigt, sondern auch, allerdings mit geringerer Gewichtung, der „Impact“, welcher die Forschung u.a. für Gesellschaft, Wirtschaft, Umwelt im Sinne von Transfer einnimmt (25%), sowie das „scientific environment“ (15%), also die Arbeits- und Forschungsbedingungen am Department (UK Research and Innovation (UKRI) 2022, S. 4). Beide gelten allerdings als schlecht messbar, manipulationsanfällig (Marginson 2015, S. 12) und als wenig aussagekräftig, da hoch positiv korrelierend mit dem Output (Pinar/Horne 2022).

⁷ Das UKRI zeichnet verantwortlich für die Durchführung des REF und ist ein Zusammenschluss von insgesamt sieben fachgruppenspezifischen Research Councils sowie des Innovate UK und Research England (<https://www.ukri.org/councils/>).

⁸ Die Mittel, die entlang des REF-Rankings vergeben werden, bilden damit nur etwa ein Fünftel bis ein Viertel der Gesamtzuwendung von öffentlicher Hand für Forschung und nur etwa 4% des Geldes, was in Höhere Bildung fließt (Shackleton/Booth 2015, S. 1). Andererseits stellt die durch REF vergebene Summe für England dort zugleich „über 80% der kompetitiv vergebenen Fördergelder“ dar (Technopolis Group 2018, zit. n. Hamann 2021, S. 85).

6.2 Zielsetzungen des REF

Im Folgenden werden zunächst die Zielsetzungen des REF skizziert werden, um im Anschluss auf dieser Grundlage der Frage nach intendierten und nichtintendierten Effekten nachzugehen. Von Interesse ist hierbei nicht zuletzt die Veränderung der Ziele im Zeitvergleich.

Das ursprüngliche Ziel („rewarding excellence“ (Stern 2016, S. 49), exzellente Forschung zu identifizieren und gezielt durch nationale Mittel zu fördern) ist nach wie vor zentral. Gleichzeitig zeigt sich bereits in frühen Stadien des REF, dass die Zielstellungen durchaus divers sind:

They (the 4 funding bodies) aim to support “a dynamic and internationally competitive UK research sector that makes a major contribution to economic prosperity, national wellbeing and the expansion and dissemination of knowledge (REF 2014: Key facts, zit. n Stern 2016, S. 8, Anm. durch uns, US & AD).

Betrachtet man die Zielsetzungen im Zeitverlauf, so zeigen sich weitere Anpassungen. So wird als Ziel des letzten REF ausgegeben “to secure the continuation of a world-class, dynamic and responsive research base across the full academic spectrum within UK higher education” (UK Research and Innovation (UKRI) 2021).

Während die o.g. Zielsetzungen in sehr allgemeiner Form die Motivation zur Implementierung des REF beschreiben, zeigen sich in der Konkretion weitere Zwecke, die unterschiedliche Aspekte adressieren und sich grob in der Weise kategorisieren lassen, dass auf der einen Seite Wettbewerb, verbunden mit indirekten Ausstrahlungseffekten, und auf der anderen Seite die strategische Ausrichtung auf nationaler und Ebene der HEI sowie Reputationseffekte gefördert werden sollen, wobei diese Differenzierung nicht überschneidungsfrei ist.

a) Förderung von Wettbewerb

Die starke Wettbewerbsorientierung soll zunächst sowohl dazu führen, international vergleichend erfolgreich zu sein als auch zur politischen Legitimation für finanzielle Zuwendungen beitragen (UK Research and Innovation (UKRI) 2022; Stern 2016, S. 9).

- Das REF führt in erster Linie zur Ausschüttung von nationalen Forschungsgeldern an die jeweiligen Departements (die diese dann übrigens nach Gusto verausgaben können, d.h. nicht nur an dasjenige wissenschaftliche Personal, deren Outputs prämiert wurden (Hamann 2021, S. 85; UK Research and Innovation (UKRI) 2021)). Grundlage ist hierbei die Konzentration auf die besten Wissenschaftler:innen des Departments. Damit verbunden ist der Effekt, dass Departments darum bemüht sind, exzellente Wissenschaftler:innen anzulocken und mithin die Konkurrenz um ‚die besten Köpfe‘ zu befördern.

- Die Ergebnisse des REF sollen jedoch nicht nur die besten Wissenschaftler:innen, sondern gleichfalls die besten Studierenden anlocken (UK Research and Innovation (UKRI) 2022, S. 3; Marginson 2015, S. 12).
- Darüber hinaus soll es weitere Geldgeber (aus Wirtschaft oder Stiftungen) motivieren, eben dort ihr Geld anzulegen, wo erfolgreiche HEI bzw. Departments durch das REF identifiziert werden konnten (UK Research and Innovation (UKRI) 2021; Marginson 2015, S. 12).
- Und schließlich ist absurderweise einer der genannten Zwecke des REF, dass man sich besser auf das kommende REF vorbereiten könne (UK Research and Innovation (UKRI) 2021, S. 3).

b) Strategische Ausrichtung und Reputation

- Durch das Verfahren soll zudem die Strategiebildung in den HEI sowie auf nationaler Ebene forciert werden (UK Research and Innovation (UKRI) 2021), um Prioritäten in verschiedenen Wissenschaftsgebieten zu setzen (Stern 2016, S. 9).
- Damit verbunden sein soll auch die Sichtbarkeit über die nationalen Grenzen hinaus, indem bspw. die häufig herausragenden Bewertungen als „world-leading“ Exzellenz die Außenwahrnehmung der UK-Universitätslandschaft sich auf diejenigen HEI konzentriert, die im REF die obersten Plätze belegen („*showcasing UK-research before the world*“) (Marginson 2015, S. 12; Hamann 2021, S. 85). So dienen die Ergebnisse des REF den jeweiligen HEI als kommunikatives Aushängeschild (UK Research and Innovation (UKRI) 2022, S. 3).
- Als weiter strategischer Effekt lässt sich schließlich die Etablierung von Standards nennen im Sinne eines „Benchmarking and establishing reputational yardsticks with peer institutions and departments“ (UK Research and Innovation (UKRI) 2022, S. 3).

Aufgrund der vielfachen Kritik seit Einführung des RAE/REF (z.B. Münch/Schäfer 2014; Marginson 2015; Shackleton/Booth 2015; Bishop 2016; Marenbon 2018; Marques et al. 2018; UCU | University and College Union 2021), wurden die Bewerbungsbedingungen, Messverfahren und Benotungsgrundlagen stetig nachjustiert (Stern 2016; Hamann 2021, S. 84). In den hier aufgeführten Beispielen für nichtintendierte Effekte werden Ziele, Bemessungen und Effekte aus unterschiedlichen RAE/REF Durchläufen exemplarisch herausgegriffen.

6.3 Intendierte und nichtintendierte Effekte des REF

Zunächst kommt man nicht umhin, dem Forschungsassessment großen Erfolg bzgl. der ursprünglichen Absicht zuzugestehen: Die Spitze (=in der Vergangenheit exzellentesten Research-Outputs vorweisend) der UK-Forschung zu identifizieren und damit deren künftige finanzielle Versorgung zu sichern (Stern 2016, S. 9). Außerdem wird, wie beabsichtigt, Forschung stärker in den Fokus gerückt, der

akademische Arbeitsmarkt beeinflusst und exzellente Forschungsleistung für den Wettbewerb gefördert (Marginson 2015, S. 13; Stern 2016, S. 9).⁹

In dem von Stern durchgeführten Survey wurde zudem angeführt, dass die Einführung von RAE/REF durchaus die erwünschten positiven Effekte für Universitäten auf Ebene von Strategiebildung und Außendarstellung hätte: diese würden durch das Assessment ein Bild ihres eigenen Profils erhalten, könnten daran strategisch anknüpfen und selbiges u.a. für Marketingzwecke nutzen (Stern 2016, S. 9). Auch für die Scientific Community sei es sinnvoll, indem es zur Ausbildung und Reflexion fachlicher Standards beitrage: „REF has helped to develop research norms and expectations at a discipline level” (Stern 2016, S. 9).

Allerdings lassen sich auch Effekte beobachten, die im Sinne Mertons in unterschiedlicher Weise als nichtintendiert klassifiziert werden können.

a) Kosten

So zeigt sich, dass nicht zuletzt aufgrund der Nachjustierungen, die im Laufe seines Bestehens am REF vorgenommen wurden, dieses stetig komplexer, aufwendiger und schließlich auch teurer wurde (Stern 2016, S. 10). Die Kosten des für 2014 durchgeführten Assessments werden auf 246 Mio. £ veranschlagt, wobei geschätzt wird, dass die tatsächlichen Kosten, die u.a. an den begutachteten Universitäten selbst entstanden sind, weit darüber hinausgingen (Shackleton/Booth 2015, S. 3). Angesichts der ursprünglichen Motivation, die begrenzt zur Verfügung stehenden Forschungsgelder effizient einzusetzen, erscheint Kostenzunahme für den administrativen Aufwand im Sinne Mertons geradezu als ein (langfristig zwar) pervernes Resultat – dies umso mehr, als der Anteil der über das Ranking vergebenen Gelder nur etwa 4% der insgesamt an die Universitäten verteilten Mittel ausmache (Shackleton/Booth 2015, S. 2). Die Autoren urteilen entsprechend: „The impact of the REF on higher education institutions is probably one of the best examples in public finance of a tail wagging a dog” (Shackleton/Booth 2015, S. 3).

b) Vorselektion durch Aufwand

Die eigentliche Zielsetzung, hohe Qualität zu identifizieren, wird – so zeigen die Effekte des REF – insofern ‚gebrochen‘, als die für die Universitäten entstehenden Kosten selbst zu einer gewissen Vorselektion führen. Der finanzielle Einsatz zur Vorbereitung und Durchführung des REF wird zum größten Teil von den teilneh-

⁹ “Much UK research is world-leading and the trend in quality continues in an upward direction, in part driven by and dependent upon the selective allocation of funding. This is demonstrated through external indicators such as citation information, where the UK share of articles in the top 1 percent of cited papers has increased from 11% in 1996 to 16% in 2012” (Stern 2016, S. 9). Allerdings gelangen Münch und Schäfer im nationalen Vergleich zu etwas anderen Einschätzungen: “This comparison of Germany, the UK and the US teaches us, that more than 25 years of competitive resource allocation and its translation into a clear-cut ranking of universities leading to a highly stratified system did not advance British science compared to Germany with its still much less stratified system, the competitive allocation of resources and rankings still in their infancy” (Münch/Schäfer 2014, S. 73).

menden HEI selbst getragen (Stern 2016, S. 9) und diese Kosten steigen in der Tendenz (Stern 2016, S. 10-11). Die enorme bürokratische und monetäre Belastung (Shackleton/Booth 2015, S. 1; Marenbon 2018) hat zur Folge, dass sich die Teilnahme an dem Wettbewerb, verbunden mit dem Risiko, im Ranking nicht gut platziert zu sein und also nicht gefördert zu werden, nicht als lohnenswert für kleinere Universitäten darstellt (Shackleton/Booth 2015, S. 3), wohingegen größere HEI mit höheren Budgets durch dieses Szenario weniger bedroht sind.

Insofern lassen sich die damit verbundenen Effekte durchaus als nichtintendiert einordnen, als die Motivation für Forschung auf systemischer Ebene kaum gefördert wird, sondern in der Kosten-Nutzen-Abwägung der Invest in Forschung nur für diejenigen lohnend erscheint, die per se in hohem Maße an finanziellen Allokationen partizipieren.

c) Matthäus-Effekt

Es ist allerdings nicht nur die finanzielle Kraft der jeweiligen Universität, die eine (erfolgreiche) Teilnahme am REF fördern oder verhindern kann: große, etablierte¹⁰ Universitäten, die drittmittelstark¹¹ sind, so zeigen die Erfahrungen, werden in der Regel durch das REF-Ranking favorisiert (Marques et al. 2018), was dazu führt, dass die bestehenden Differenzierungen innerhalb der Forschungslandschaft fortgeschrieben werden bzw. weiter zunehmen und die REF-basierte Förderung bei den gleichen wenigen¹² ankommt, die zuvor schon gut aufgestellt waren (Shackleton/Booth 2015, S. 4). Auch wenn diese Art der Vergabe durchaus mit dem Ziel, Exzellenz konzentriert zu fördern, konform ist, muss angezweifelt werden, dass eine Zementierung der Aufteilung in „oben“ und „unten“ oder ein weiteres Auseinanderdriften aufgrund des Matthäus-Effekts als absichtsvoll gelten kann. Sich-selbst-verstärkende Prozesse sind nicht inhärent alleine bedingt durch das REF, noch sind sie durch andere Assessmentformen immer ausgeschlossen – die Art, wie das REF misst, identifiziert jedoch weniger Forschungspotentiale als exzellente, weil bereits geförderte Arbeit.

d) Noteninflation

Auch die zuvor angeführte Beobachtung, dass das REF zu einer hohen Anzahl an ‚world-leading‘-Forschung in UK führt, legt zwar einerseits nahe, dass das Ziel, zu einer hohen Reputation nationaler Forschung beizutragen, zunehmend erfüllt wird, widerspricht allerdings dem Anspruch einer Bewertungspraxis, die Diffe-

¹⁰ Als eine weitere Ursache, für die sich manifestierende Konzentration auf dieselben Einrichtungen, vermutet Marginson den „Halo-Effekt“ – die eine Eigenschaft (nämlich, dass der eingereichte Forschungoutput z.B. vom renommierten Imperial-College stammt) beeinflusst die Zuschreibung von anderen Eigenschaften (also die Bewertung) (Marginson 2015, S. 13).

¹¹ Marques/Powell/Zapp/Biesta zeigen, dass zwischen 2001 und 2014 das REF-Ranking mehr und mehr mit den Drittmittelzuwendungen korreliert (Marques et al. 2018).

¹² Es nimmt kaum Wunder, dass es „Oxbridge“, das University College London und weitere aus der „Russell-Group“ sind, die zwar nur einen kleinen Teil (15%) der UK-Universitäten darstellen, aber fast 70% der Forschung einreichen, die vom REF als „world-leading“ taxiert wurden (Shackleton/Booth 2015, S. 4).

renzen identifiziert und daran anschließende strategische Entscheidungen auf organisatorischer und systemischer Ebene vorzubereiten hilft.

e) Forschungsausrichtung und Publikationsverhalten

Aber nicht nur die Bewertungsskalen, sondern auch die Methode hat unerwünschte Nebeneffekte: Stern fand heraus, dass die Einteilung in disziplinäre Units-of-Assessment das Einreichen und vermutlich im Vorfeld auch das Produzieren von interdisziplinären Forschungen tendenziell hemmt (Stern 2016, S. 13-15). Dies gilt sowohl für Forschungspublikationen als auch für die REF-Einreichung von Impact-case-studies (Stern 2016, S. 17), die interdisziplinär, aus Kollaborationen verschiedener Universitäten oder international entstanden sind.

Die Anforderung, eine fixe Anzahl (vier) an Publikationen pro Person im Bemessungszeitraum zur Bewertung zu nominieren, befördert zudem das zuvor beschriebene publikatorische ‘Salami-Slicing’ von Forschungsergebnissen (Shackleton/Booth 2015, S. 3; Stern 2016, S. 15)¹³: Statt eines Buches werden tendenziell mehrere Artikel¹⁴ veröffentlicht. So resümiert Hamann, dass diese Praxis entlang des REF auch mehr und mehr zur Devise in Disziplinen wie der Geschichtswissenschaft, deren übliche renommierte Veröffentlichungsform das Buch war, wird. Zum gleichen Ergebnis gelangen fachübergreifend auch Marques/ Powell/Zapp/Biesta:

We believe this to be a largely unintended consequence of the exercise; although there is no official requirement in the research evaluation system that privileges the research article as the preferred format of output, the bulk of all submissions are articles (Marques et al. 2018).

Wenngleich das REF in seiner grundlegenden Ausrichtung hierzu keine Zielsetzungen formuliert, trägt damit seine Bewertungspraxis offenbar dazu bei, dass eine an den MINT-fächern orientierte Forschungs- und Publikationsausrichtung dominiert und einen zunehmenden und in diesem Fall zumindest nicht explizit intendierten Effekt aufweist:

The criteria [Impact] not designed for humanities subjects, but rather for scientific discoveries or technological advances. They exclude academic books that reach a wide audience of general readers – the real way scholars in history, literature, art and philosophy make an impact (Marenbon 2018, Ergänzung durch US & AD).

¹³ Marques, Powell, Zapp & Biesta haben festgestellt, dass zwischen 2001 und 2014 die Anzahl aller eingereichten anderen Publikationsformen zu Gunsten des Artikels geringer wurden (Marques et al. 2018).

¹⁴ Allerdings, so führt Marenbon an, könne eine Monographie im REF auch manchmal als zwei Publikationen gewertet werden, auch das sei jedoch unverhältnismäßig, da Monographien „might be 20 times the length of an article and require 50 times as much work“ (Marenbon 2018).

Dies betrifft auch die Idee und das Verfahren der Impact-Messung: Der Transfer und Benefit ökonomischer oder sozialer Art ist im Kontext geisteswissenschaftlicher Forschung häufig schwieriger messbar, verwendet man nicht aufwendige Verfahren, die auch langfristige Wirkungen mit in den Blick nehmen. Zumal es in den Geisteswissenschaften grundsätzlich umstritten ist, inwieweit Anwendungsorientierung forschungsrichtungslenkend sein sollte.

Schließlich muss auch der Befürchtung Beachtung geschenkt werden, dass die zeitplanerische Orientierung am REF und die „Belohnung von Erfolg“, die Aufnahme von langfristigen, profunden und potenziell riskanten Forschungsvorhaben verhindere oder zumindest beschränke (s. auch Shackleton/Booth 2015, S. 1; Marenbon 2018):

This can drive them towards safe topics and short-termism, and a reluctance to engage in risky or multidisciplinary projects, in order to ensure reliable, high quality publication within the REF period, and may be discouraging innovative thinking and risk taking (Stern 2016, S. 14).

Es kann daher geschlossen werden, dass die REF-Bewertung dazu beiträgt, dass die geistes- und sozialwissenschaftlichen Disziplinen sich in Forschungsdesign, Publikationskultur und Wissenschaftstransfer dem Verhalten der MINT-Fächer annähern (Marques et al. 2018; Hamann 2021, S. 88). Hamann betont, dass es sich bei den Publikationsaktivitäten nicht um das „generelle Auflösen von Fachkulturen“ handle, sondern vielmehr um eine „Entdifferenzierung disziplinärer Kulturen“ (Hamann 2021, S. 83). Marenbon gelangt entsprechend zu dem Ergebnis: „for the humanities, the Ref does nothing but harm“ (Marenbon 2018).

f) Wilderungsverhalten

Die Bemessung der Forschungsleistung eines Departments wird, wie oben beschrieben, anhand der Einreichungen der Wissenschaftler:innen vorgenommen, die das Department „ins Rennen schickt“, mit dem Ergebnis, dass

Across the UK, of 194,000 academic staff in post in 2013-14, only 52,000 were submitted to REF 2014. Within some universities this selectivity can be carried to absurd lengths, with cases of departments with more than a hundred staff being represented by fewer than ten individual submissions (Shackleton/Booth 2015, S. 3).

Marques/Powell/Zapp/Biesta stellen fest, dass der Anteil der REF-einreichenden Wissenschaftler:innen pro Departement konstant abnimmt und werten dies als Ergebnis strategischer Entscheidung der Departments (Marenbon 2018). Das Zurückbleiben und vom REF ausgeschlossen zu werden (u.a. auch, weil man auf weniger als 4 Publikationen im relevanten Zeitraum kommt), führ(t)(e) zur Demoralisierung von Forscher:innen (UCU | University and College Union 2021)

Ob also eine zutreffende Aussage über die Gesamtforschungsstärke eines Departments oder von Universitäten ermittelt werden kann, ist anzuzweifeln

(Shackleton/Booth 2015, S. 3). Es profitieren gleichwohl auch die weniger exzellenten Forscher:innen eines großen, erfolgreichen Departments, wohingegen ähnlich talentierte in kleineren HEI ohne "Zugpferde" keine Chance auf QR-Fundings haben. Die Bedeutung dieser Spitzen-Forscher:innen für das Department und die Gesamt-HEI hat in der Vergangenheit zu einem Head-Hunting seitens der Universitäten insbesondere kurz vor dem Zensus-Datum geführt (Stern 2016, S. 13). Die REF-Relevanz einzelner Wissenschaftler:innen bedingte zudem insgesamt eine wachsende Ungleichheit der Gehälter (Shackleton und Booth 2015, S. 1; Stern 2016, S. 12). Schließlich würden eingedenk dessen die HEI nun zurückhaltender, langfristige Forschungen zu finanzieren aus Befürchtung, die dadurch ermöglichten Outputs könnten bei Weggang des Personals anderen HEI zugerechnet werden.

Diese Konkurrenz der Universitäten untereinander um ein limitiertes Gut (=forschungsstarke Wissenschaftler:innen) kann aber aus der Logik heraus nicht dem Voranschreiten der nationalen Forschungsstärke dienen: „this may boost one institution’s score at the expense of another, but does nothing to boost the sector’s total research output“ (Shackleton/Booth 2015, S. 3).

g) Vernachlässigung der Lehre

Einer der größten Kritikpunkte, die dem RAE/REF gegenüber angebracht wurden, ist die Lenkung der Aufmerksamkeit hin zur Forschung und weg von Lehre (UCU | University and College Union 2021; Shackleton/Booth 2015, S. 1, 3; Stern 2016, S. 16; Bishop 2016, S. 16). Wissensvermittlung würde zunehmend von Promovierenden übernommen und auch andere akademische oder universitäre Aktivitäten würden zugunsten der Forschung reduziert (Shackleton/Booth 2015, S. 3).

Wohl auch auf Sterns Umfrage hin wurde 2017 analog zum REF ein TEF (Teaching Excellence and Student Outcomes Framework) vom Department of Education ins Leben gerufen, welches englische Universitäten hinsichtlich der Lehre bewertet (Indikatoren: Studienabbruchrate, Studierendenzufriedenheit, Vermittlung in den Arbeitsmarkt nach Studienabschluss). Auch beim TEF gibt es eine Verbindung zur jeweiligen Finanzierung der HEI: je nach Ranking dürfen diese ihre Studiengebühren (gemessen an der Inflation) erhöhen (Adams 2016). Interpretiert man die Gründung des TEF und die damit verbundene Förderung von qualitativ hochwertiger Wissensvermittlung als REF-Effekt ein, so ließe er sich als zwar nichtintendiert, aber durchaus positiv charakterisieren.

7. Fazit

Fasst man die vorherigen Ausführungen zu den Effekten der Forschungsbewertung mit Blick auf die Kategorisierung nach Merton zusammen, so zeigen sich in allen vier Kategorien entlang der Dimensionen latent vs. manifest und funktional vs. dysfunktional (vgl. Abb. 1) Wirkungen, die von unterschiedlichen Bewertungspraktiken ausgehen: Es lassen sich sowohl – wie beim REF erläutert – die

Realisierung intendierter Effekte als auch perverse Resultate, unerwarteter Rückschläge und – in vielleicht geringerem Maße – glückliche Zufälle ausmachen.

Bemüht man die Überlegungen Mertons zu Ursachen auftretender nichtintendierter Effekte, so lässt sich zunächst festhalten, dass unvollständiges Wissen der Akteure eine wesentliche Erklärung bietet. So zeigen die Anpassungen der Bewertungsmodi im REF, aber auch die Bemühungen um eine Anpassung der Verfahren im Kontext von Rankings und Ratings, dass Forschungsbewertung auf keine als ‚stabil‘ wahrgenommenen Indikatoren zurückgreifen kann.

In gleichem Maße sind damit Fehleinschätzungen verbunden, wie sich sowohl im Rahmen des REF, aber auch darüber hinaus zeigt, indem z.B. keine hinreichende Differenzierung zwischen Geistes- und Sozial- sowie naturwissenschaftlichen Forschungsmodi vorgenommen wird. Die Fokussierung auf Artikel in Fachzeitschriften und die Orientierung am Journal Impact führt in diesem Zusammenhang gewissermaßen zu einer „Vernaturwissenschaftlichung“ der Forschungsbewertung.

Auch die von Merton angeführten Ursachen für nichtintendierte Effekte der Unmittelbarkeit des Interesses sowie der grundlegenden Werte lassen sich für die hier betrachteten Verfahren belegen, indem insbesondere die im REF zu beobachtende wachsende Differenz zwischen dem Engagement in der Forschung oder in der Lehre der dominanten Fokussierung auf Forschungsexzellenz zuzuschreiben ist. Die Ergänzung um das in UK eingeführte Ranking in der Lehre lässt sich als Indiz dafür interpretieren, dass zwar eine Priorisierung der Forschung durchaus gewollt war, der damit verbundene Effekt auf das Engagement in der Lehre jedoch nichtintendiert und im Sinne Mertons latent und dysfunktional war.

Ungeachtet dieser Ergebnisse lässt sich insgesamt festhalten, dass Forschungsbewertung – so auch durch das REF – eigene Handlungslogiken und gewissermaßen eine Professionalisierung im Umgang mit Erfolgskriterien und Bewertungsverfahren evoziert:

These results demonstrate strong “reactivity” that we framed as “reverse engineering”, in which, over time, managers and academics learnt how to “play the game” in order to maximise performance and their departmental results in the exercise, with less regard for the individual or disciplinary levels (Marques et al. 2018).

Literaturverzeichnis

Adams, R. (2016): English universities to be ranked gold, silver and bronze. Ratings form part of government’s teaching excellence framework and will come into force from the middle of next year. The Guardian Online. <https://www.theguardian.com/education/2016/sep/29/english-universities-ranked-gold-silver-and-bronze> (27.05.2023).

- Bishop, D. (2016): Clarity of purpose in the TEF and the REF. Ministers must demonstrate that assessing 'excellence' does more good than harm, says Dorothy Bishop. Times Higher Education.
<https://www.timeshighereducation.com/blog/clarity-purpose-tef-and-ref>, (05.11.2022).
- Hamann, J. (2021): Disziplinäre Entdifferenzierung durch Leistungsbewertung? Der Fall Geschichtswissenschaft in Großbritannien. In: Das Hochschulwesen (HSW), 69 (3+4), S. 83-89. DOI: 10.31235/osf.io/8mj67.
- Hönig, B. (2020): Nicht-intendierte Folgen absichtsgeleiteten Handelns. In: Fleck, C./Dayé, C. (Hg.): Meilensteine der Soziologie. Unter Mitarbeit von M. Duller. Frankfurt/New York, S. 305-311.
- Hornbostel, S. (1997): Wissenschaftsindikatoren. Bewertungen in der Wissenschaft. Opladen.
- Hüther, O./Krücken, G. (2016): Hochschulen. Wiesbaden.
- Marenbon, J. (2018): Our research funding system is shortchanging the humanities. The humanities subjects do not benefit from the research excellence framework. They need a better system. The Guardian Online.
https://www.theguardian.com/education/2018/dec/05/our-research-funding-system-is-shortchanging-the-humanities?CMP=Share_AndroidApp_Tweet (10.04.2023).
- Marginson, S. (2015): UK Research Excellence: Getting Better All the Time? In: IHE, 81, pp. 12-13. DOI: 10.6017/ihe.2015.81.8732.
- Marques, M./Powell, J. J. W./Zapp, M./Biesta, G. (2018): The RAE/REF have engendered evaluation selectivity and strategic behaviour, reinforced scientific norms, and further stratified UK higher education.
<https://blogs.lse.ac.uk/impactofsocialsciences/2018/02/13/the-raeref-have-engendered-evaluation-selectivity-and-strategic-behaviour-reinforced-scientific-norms-and-further-stratified-uk-higher-education/> (10.04.2023).
- Marx, W./Bornmann, L. (2012): Der Journal Impact Factor – Ein problematischer bibliometrischer Indikator. In: Qualität in der Wissenschaft (QIW), 6 (2), S. 30-34.
- Merton, R. K. (1936): The Unanticipated Consequences of Purposive Social Action. In: American Sociological Review, 1 (6), pp. 894-904.
DOI: 10.2307/2084615.
- Merton, R. K. (Hg.) (1996): On social structure and science. Edited and with an introduction by P. Sztompka. Chicago: University of Chicago Press.
<http://catdir.loc.gov/catdir/description/uchi051/96000928.html>
- Merton, R. K. (1968): Social theory and social structure. New York: Free Press.
- Münch, R./Schäfer, L. O. (2014): Rankings, Diversity and the Power of Renewal in Science. A Comparison between Germany, the UK and the US. In: European Journal of Education, 49 (1), pp. 60-76. DOI: 10.1111/ejed.12065.
- Musselin, C. (2007): Are Universities Specific Organisations? In: Krücken, G./Kosmützky, A./Torka, M. (Hg.): Towards a multiversity? Universities between global trends and national traditions. Bielefeld, pp. 63-84.

- Pinar, M./Horne, T. J. (2022): Assessing research excellence: Evaluating the Research Excellence Framework. In: *Research Evaluation*, 31 (2), pp. 173-187. DOI: 10.1093/reseval/rvab042.
- Schmidt, U./Lenz, E. (2021): Funktionen von Evaluation im Hochschulkontext. In: *Handbuch für Qualität in Studium, Lehre und Forschung* (78), B 1.10, S. 1-22.
- Schwarz, C. (2006): Evaluation als modernes Ritual. Zur Ambivalenz gesellschaftlicher Rationalisierung am Beispiel virtueller Universitätsprojekte. Zugl.: Hannover, Univ., Diss., 2004. Hamburg/Münster.
- Shackleton, J. R./Booth, P. (2015): Abolishing the Higher Education Research Excellence Framework (the REF). <https://ssrn.com/abstract=3893744> (10.04.2023). DOI: 10.2139/ssrn.3893744.
- Sivertsen, G. (2017): Unique, but still best practice? The Research Excellence Framework (REF) from an international perspective. In: *Palgrave Commun*, 3 (1). DOI: 10.1057/palcomms.2017.78.
- Stern, N. (2016): Building on Success and Learning from Experience. An Independent Review of the Research Excellence Framework. Ed. by Department for Business, Energy & Industrial Strategy. Department for Business, Energy & Industrial Strategy. London, UK.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/541338/ind-16-9-ref-stern-review.pdf (25.03.2023).
- Stockmann, R. (Hg.) (2007): Handbuch zur Evaluation. Eine praktische Handlungsanleitung. Sozialwissenschaftliche Evaluationsforschung, Bd. 6. Münster, <https://elibrary.utb.de/doi/book/10.31244/9783830967668>.
- Stockmann, R./Meyer, W. (2014): Evaluation. Eine Einführung. 2. Auflage. Stuttgart. DOI: 10.36198/9783838585536.
- UCU | University and College Union (2021): RAE 2008. UCU repeatedly reiterated its policy of opposition to the Research Assessment Exercise (RAE) – a mechanism by which a large portion of higher education funding was allocated. <https://www.ucu.org.uk/rae2008>, zuletzt aktualisiert am 17.03.2021, (21.05.2023).
- UK Research and Innovation (UKRI) (2021): What is the REF? <https://www.ref.ac.uk/about-the-ref/what-is-the-ref/>.
- UK Research and Innovation (UKRI) (2022): REF 2021. Key Facts. https://www.ref.ac.uk/media/1848/ref2021_key_facts.pdf (06.05.2023).
- Wippler, R. (1978): Nicht-intendierte soziale Folgen individueller Handlungen. In: *Soziale Welt*, 29 (2), S. 155-179, <https://www.jstor.org/stable/40877220> (18.10.2022).