

Keine Exzellenz ohne Mittelklasse!

Kritik an Quantifizierungsdynamiken in Bewertungsprozessen aus dem Innern der Wissenschaft

Barbara Hendriks, Martin Reinhart, Cornelia Schendzielorz

Beitrag zur Ad-Hoc-Gruppe »Grenzen der Quantifizierung in der Datengesellschaft«

Einleitung

In Wissenschaft und Hochschule spielen Quantifizierungen, Vergleichs- und Kategorisierungsprozesse im Rahmen der Leistungsbewertung eine zentrale Rolle: als Beispiele dafür können Hochschulrankings, Publikationszahlen und Journal-Impact Faktoren (Lamont 2012, Fourcade, Healy 2013, Krüger, Reinhart 2016) herangezogen werden. Im Zuge des Wettbewerbs um finanzielle Ressourcen, welcher diskursiv häufig mit dem Begriff des New Public Managements in Verbindung steht, geraten derartige Evaluationsmechanismen als Elemente „numerokratischen Regierens“ (Angermüller 2011, S.177), im Sinne von daten- und zahlenbasierten Entscheidungsmechanismen, ins Visier wissenschaftspolitischer Auseinandersetzungen.

Vor diesem Hintergrund gehen wir der Frage nach, inwiefern sich in wissenschaftlichen Bewertungsverfahren im Rahmen von Publikationstätigkeit, Drittmittelakquise und Personalrekrutierung eine Dynamik entspinnt, in der qualitative Bewertungsprozesse durch quantitative überlagert und ersetzt werden. Ziel des Beitrags ist es, am Beispiel der Lebenswissenschaften, Effekte, Chancen und Grenzen der Quantifizierung auszuloten. Die These einer Quantifizierungsdynamik basiert auf der Analyse von Interviews mit an Schweizer Hochschulen arbeitenden Forscher/-innen, die im Rahmen einer qualitativen Fallstudie zur Forschungsförderung und -evaluation in den Lebenswissenschaften durchgeführt wurden (Hendriks et al. 2018, S.78–84; 88–92).

Die Lebenswissenschaften bieten sich als Untersuchungsfeld an, da sie aufgrund der biotechnologischen Entwicklung ein Wachstum der Forschung erleben. Angesichts der steigenden finanziellen Investitionen sind sie zudem mit Forderungen eines „return on investment“ (Kraft 2013) konfrontiert, das sich in Form von Forschungsoutput manifestieren soll. Die im Zuge dessen steigende Menge des Outputs, insbesondere von Publikationen, stellt auch die Evaluationsverfahren in den Lebenswissenschaften vor Herausforderungen, wenn es darum geht zu eruieren, welches Wissen wie wertvoll ist (siehe dazu Macleod et al. 2014 in der Lancet-Debatte „increasing value, reducing waste“). Dabei geraten Evaluationsverfahren in ihrer Funktion, Forschungsvorhaben zu klassifizieren und in nachvollzieh-

barer und begründbarer Weise zu selektieren, in den Blick, werden kritisiert und geraten mitunter selber auf den Prüfstand.

Zunächst erörtern wir, inwiefern der steigende Output im Bereich der Lebenswissenschaften quantifizierende Evaluationsmodi in Bewertungsverfahren derart befördert, dass sie die Bedeutung qualitativer Verfahrenselemente mindern und wie die wissenschaftlichen Akteure mit diesen Entwicklungen umgehen. Auf dieser Basis konstatieren wir eine Quantifizierungsdynamik in Bewertungsprozessen innerhalb der Lebenswissenschaften, die ungeachtet ihrer Limitationen anhält und die Fragen nach deren Funktionalität aufwirft. Im zweiten Teil diskutieren wir, inwiefern Praktiken der Soziokalkulation (Vormbusch 2012), die insbesondere in den Verfahren der Personalrekrutierung zur Anwendung kommen, als Teil der Quantifizierungsdynamik oder als Reaktion auf die Grenzen der Quantifizierung zu begreifen sind. Abschließend erörtern wir die Rechtfertigungsmuster, welche die Basis der Kritik der Akteure bilden. Dabei greifen wir auf das begriffliche Repertoire einer Soziologie der Kritik (Boltanski, Thévenot 2007) zurück. In Referenz darauf lässt sich an der Kritik der Befragten Forscher/-innen eine gesellschaftspolitische Ebene der Problematisierung von Quantifizierung aufzeigen.

Ausgangslage und methodisches Vorgehen

Bewertungs- und Peer-Review-Prozesse sind zentrale Mechanismen der Qualitätssicherung in der Wissenschaft (Reinhart 2012). Innerhalb dieser Prozesse wird mittels Verfahren und Kriterien beispielsweise über die Veröffentlichung eines Manuskripts, die Bewilligung eines Forschungsvorhabens oder in personalisierten Evaluationsverfahren über Beförderungen, Entfristungen und Stellenbesetzungen entschieden. Damit sind sie zentrale Bestandteile der Mechanismen, mit denen Wissenschaft sich selbst zu steuern beansprucht und die in unterschiedlichsten wissenschaftspolitischen Kontexten zur Begründung einer Autonomie der Wissenschaft herangezogen werden (Wilholt 2012). Als solche sind sie gleichermaßen Ansatzpunkt spezifischer Verfahrenskritik wie Gegenstand allgemeiner Formen der Wissenschaftskritik. Folglich ist zu ergründen, inwiefern Bewertungsprozesse mittelbar Einfluss auf die wissenschaftliche Praxis nehmen, das heißt, welche Effekte sie für die Karrieren einzelner sowie für die Entwicklung von wissenschaftlichen Teilbereichen zeitigen. Vor dem Hintergrund dieses Forschungsdesiderats hat der Schweizer Wissenschafts- und Innovationsrat 2017 eine empirische Studie in Auftrag gegeben, die die Effekte von Evaluations- und Finanzierungspraktiken in den Lebenswissenschaften im Kontext der Schweiz explorativ erfassen soll. In der Studie wurden 22 in Forschung, Drittmittelakquise und Begutachtung erfahrene Professor/-innen und Postdoktorand/-innen verschiedener Hochschulen befragt.¹

Bestandsaufnahme: Zentrale Bewertungsprozesse in den Lebenswissenschaften

Die zentrale Bedeutung von Bewertungs- und Peer-Review-Prozessen manifestiert sich in zu unterschiedlichen Zwecken organisierten Bewertungsprozeduren, wie sie die Verfahren zur Manuskriptbegutachtung bei Zeitschriften, zur Antragsbegutachtung bei der Drittmittelakquise und zur Personal-

¹ Das methodische Vorgehen bei Erhebung und Auswertung ist ausführlich dargelegt in Hendriks et al. 2018, S.68-71.

auswahl bei der Stellenbesetzung darstellen. Im Kern verweisen jedoch alle Verfahren auf eine als fachspezifisch verstandene Expertise, die oft mit der Bezeichnung „peer review“ assoziiert wird. Diese Verfahren enthalten immer Formen der Qualitätsbeurteilung, die sowohl als qualitativ als auch als quantitativ bezeichnet werden können, was sich auch anhand der Studie zeigen lässt.

Mit Blick auf die Publikationstätigkeit in den Lebenswissenschaften, stellt das *Journal Peer Review* ein zentrales Bewertungsverfahren dar. Es ist von qualitativen Elementen geprägt, denn es wird auf der Grundlage von Gutachten und Voten der Herausgeber/-innen die Bedeutung und Ergiebigkeit der dargelegten Arbeiten für das Forschungsfeld bewertet und mit der Veröffentlichung oder Nicht-Veröffentlichung von Manuskripten über die Mitteilungswürdigkeit der Forschung entschieden. Darüber hinaus betonen die Befragten der Studie einhellig, dass im Bereich der Lebenswissenschaften in allen Evaluationsverfahren die Publikationsliste und zudem die Gewichtung der Publikationen anhand von Metriken, wie Zitationsindizes und/oder Journal Impact Faktoren ein gewichtiger Referenzpunkt sind. Somit nimmt auch der quantifizierte und messbare Publikationsoutput in den Lebenswissenschaften einen zentralen Stellenwert in der Bewertung wissenschaftlicher Leistung ein.

Bei der *Drittmittelvergabe* werden in der Regel das jeweilige Antragsdokument und etwaige Präsentationen hinsichtlich der wissenschaftlichen Qualität, der Relevanz des Vorhabens, der Originalität, des Innovationspotenzials etc. in Form von Gutachten im Fließtext zunächst *qualitativ* bewertet. Zudem fließen die bisherigen Erfolge in der Mittelakquise und der Forschungserfahrung anhand der im Lebenslauf gelisteten eingeworbenen Beträge, Preise und Auszeichnungen sowie weitere Arbeitsergebnisse auf der Grundlage von vorangegangenen Projekten in die Bewertung mit ein. Dieser sogenannte bisherige Forschungsoutput wird wiederum häufig quantifiziert erfasst; konkret berichten die befragten Lebenswissenschaftler/-innen, dass die Menge der eingeworbenen Forschungsförderungen, der Publikationen, der Patente, der Spin-offs oder Ausgründungen gezählt und der Impact dieser Outputs in Zitationsindizes, Journal Impact Faktoren oder aber in Patenten und deren Nutzungen etc. gemessen wird.

Auch die Bewertungsverfahren der *Personalrekrutierung* verbinden qualitative Elemente, wie Anschreiben, Vorstellungsgespräche etc. mit quantitativen Elementen (Publikationslisten etc.). Die befragten Lebenswissenschaftler/-innen äußern, dass besonders bei längerfristigen Stellenbesetzungen, -beförderungen und -entfristungen sowie bei der Verteilung interner Mittel vermehrt Evaluationsverfahren zur Anwendung kommen, in denen auf der Basis der Menge von Publikationen und der eingeworbenen Drittmittel vergleichend bewertet wird. Hinzu kommen für die Lebenswissenschaften häufig Gewichtungen des Outputs anhand von Impact Faktoren und Zitationsindizes, wodurch zusätzlich Impact Faktoren zum zentralen Bewertungskriterium werden.

Die Studie zeigt ferner, wie sich die Anforderungen an das wissenschaftliche Personal im Bereich der Lebenswissenschaften entlang der steigenden Karrierepositionen akkumulieren und sogenannte *hard facts* an Bedeutung gewinnen.

Qualitative Elemente, wie beispielsweise Empfehlungsschreiben sowie Motivation und Forschungserfahrung, werden durch quantitative Elemente, wie Publikationsoutput und Journal Impact Factor, Zitationsindizes, Anzahl und Höhe der eingeworbenen Drittmittel, komplementiert und überlagert (Hendriks et al. 2018, S.84). Empirisch zeigt sich, dass sich die Funktionslogik der Personalrekrutierung wandelt: und zwar von einem System des kollegialen Vertrauens, in dem erfahrungsbasierte, qualitative, subjektive Einschätzungen ausgetauscht werden (wie bei der Rekrutierung von Doktorand/-innen) hin zu einer Selektion entlang von ‚harten‘ Kriterien, die zählbar und gegebenenfalls messbar sind und somit quantifiziert und teilweise auch kalkulatativ erfasst werden können (beispielsweise bei der Vergabe von Leitungsstellen und entfristeten Posten, beispielsweise den Professuren). Die empirischen Daten zeigen, dass der Versuch der quantifizierenden Repräsentation erkennbar mit fortschreitender

Karriere und höherer beruflicher Position zunimmt, wobei gleichwohl eine Verzahnung mit qualitativen Verfahren bestehen bleibt.

Die Kritik der wissenschaftlichen Akteure

Auf Basis der Fallstudie lässt sich mit Blick auf die Evaluation von Forschungsleistungen in den Lebenswissenschaften mit fortschreitender Karriere eine Verschiebung im Verhältnis qualitativer und quantitativer Bewertungselemente zu Gunsten einer stärkeren Gewichtung quantifizierter Bewertungsgrundlagen beobachten. Vor dem Hintergrund der geltenden Reputationsordnung (Knie, Simon 2016, S.31) zeichnet sich somit eine Dynamik der quantifizierten Leistungsbewertung ab, die tendenziell mit steigender Position an Bedeutung gewinnt, da sie in Selektionsprozessen als Kontextinformationen zur Klassifikation und Einordnung der Projekte bzw. Kandidat/-innen genutzt wird. Ferner zeigen die empirischen Daten, dass die Quantifizierung in Verbindung mit qualitativen Verfahrenselementen von den Befragten differenziert betrachtet wird. Neben der Reduktion an Information im Zuge der Quantifizierung und standardisierender Vergleichsmaßstäbe wird ebenso das produktive bzw. funktionale Element einer quantifizierten Erfassung wissenschaftlicher Leistungen gesehen und in den spezifischen Bewertungsverfahren entsprechend gewichtet.

Die kritische Reflexion einer Quantifizierung zeigt sich im Material insbesondere daran, dass die zunehmende Bedeutung von zitationsbasierten und am Publikationsort festgemachten Metriken in der Urteilsbildung unter den Befragten umstritten ist. Publikationen gelten im akademischen System zwar als zentrale Währung (Hornbostel 2015) und sind als Arbeits- und Leistungsnachweise anerkannt. Viele der befragten Wissenschaftler/-innen kritisieren jedoch die Dominanz der Quantität von Publikationen bei der Verwendung dieser Metriken: insbesondere wenn diese im Kontext von Drittmittelakquise und Stellenbesetzung, der Vergabe institutioneller-finanzieller Ressourcen (beispielsweise LOM), interner Beförderung oder Entfristung bedeutsam werden. Diese Praxis der Leistungsbewertung wird sodann als eine inadäquate Reduktion auf scheinbar objektiv vergleichbare Zahlen gedeutet.

„What I'm very worried about is the sort of coming down to certain factors, whether it's a citation index and a multiplication, or H-factor, [...] because all these numbers have inherent flaws, and some more bigger problems than others“ (Herr Friedhelm, Prof., Abs. 27).

„I'm very much against the quantitative methods for evaluating research because they are based on the impact factors of journals, which is bullshit. Therefore, there is no way of quantitatively evaluating the discovery that was done one year ago. It should be based on the honest opinion of peers or people from the field, how much this work advanced the research. So the number of publications means nothing because now everyone is playing the game, right? The more you publish, the better you're funded. And the impact factor of publication means little until you read the paper and read and revise what was done“ (Frau Todorova, Prof., Abs. 67).

Zusammenfassend ist festzustellen, dass die Befragten die zunehmende Gewichtung quantifizierter Bewertungselemente zumindest als ambivalent erachten und diesen negative Auswirkungen auf die wissenschaftliche Arbeitspraxis zuschreiben.

Die Befragten äußern darüber hinaus die Befürchtung, dass die Entscheidung, was, in welcher Form, wo publiziert wird, sich nicht daran orientiert, welche Forschungen mit Blick auf die Disziplin

und deren Erkenntnisfortschritt für wen relevant und mitteilenswert ist; sondern, dass sich die Publikationstätigkeit primär danach richtet, „diesen Parameter hochzuschrauben“ (Herr von Stade, Ass-Prof., Abs. 109), also danach, welche Art der Veröffentlichung der eigenen Förderfähigkeit und Karriere am zuträglichsten ist.

In dieser Kritik scheint eine *Dynamik* in der Leistungsmessung und -evaluation auf, welche die Referenz auf quantifizierte Bewertungsparameter und deren Weiterentwicklung befördert. Wenn durch das Zählen von Publikationen in erster Linie die Quantität im Forschungsoutput stetig belohnt wird, regt es weitere Publikationstätigkeiten in hoher Frequenz an. Das wiederum erzeugt den Bedarf nach einer Klassifizierung und einem Ranking von Publikationen, beispielsweise mit Hilfe von Indikatoren wie Zitationsindizes, Journal Impact Faktoren, Journal Rankings etc. Denn letztlich versprechen diese Indizes eine Art Hilfestellung zu leisten, um ‚unwichtigen‘ von ‚wichtigen‘ Artikeln zu unterscheiden; und zwar bevor man sie alle gründlich gelesen hat, weil dieses im Zuge der zunehmenden Publikationszahlen kaum noch zu bewältigen scheint. Damit leisten Indizes eine konkrete Hilfestellung in Form einer Vorselektion zwischen lesenswerten und nicht lesenswerten Beiträgen.

Da die Befragten als Professor/-innen und Postdoktorand/-innen selber auf vielfältige Weise in die von ihnen kritisch betrachteten Bewertungsprozesse eingebunden sind, stellt sich die Frage, *wie* die Akteure in ihrer wissenschaftlichen Arbeitspraxis mit den von ihnen kritisierten Bewertungsprozessen umgehen?

Hier zeigt sich, dass die Befragten reflektieren und sich darüber bewusst sind, dass die Vergleichbarkeit der Zahlenwerte von Metriken und Scores, die zur Leistungsbewertung genutzt werden, nur scheinbar Objektivität verbürgt. Die Reduktion der Sachverhalte im Zuge der Quantifizierung wird von vielen als inadäquat dargestellt.

“Because how evaluation is done now is based on the number of publications and impact factor of publications. So, many people equal discovery to publication in Nature but this is not true” (Frau Todorova, Prof., Abs. 41).

Trotz der Kritik werden diese Bewertungsprozeduren auch von den hier zitierten Kritiker/-innen, konkret den befragten Lebenswissenschaftler/-innen, praktiziert. Postdoktorand/-innen oder Assistant-Professor/-innen ohne Tenure führen vor diesem Hintergrund an, dass es für sie quasi alternativlos sei, sich innerhalb der gegebenen Bewertungsstrukturen zu profilieren und ‚mitzuspielen‘, um sich im Wissenschaftssystem eine Zukunft offen zu halten. Tiefgreifende Strukturreformen könnten laut der Befragten somit nur im Zusammenschluss jener erfolgen, welche im selben System bereits erfolgreich waren.

Gleichwohl ist unter den Befragten auch ein Bewusstsein darüber vorhanden, dass die anhaltende Verwendung, im Sinne einer alltagspraktischen Selbstverständlichkeit, zusätzlich zu einer schleichenden Akzeptanz der quantitativen Bewertungspraktiken und zu deren Wirkungsmächtigkeit beitragen kann. Angesichts der anhaltenden Verwendung derartiger Metriken ist zu erörtern, inwieweit die quantifizierenden Elemente der Bewertungsprozesse für die beteiligten individuellen und organisationalen Akteure (Förderorganisationen, Universitäten, Verlage) *funktional* und als Entscheidungsmechanismus *effektiv* sind?

Es fällt auf, dass der Einsatz von zitationsbasierten und am Publikationsort festgemachten Metriken einerseits in der Urteilsbildung bei der Drittmittel- und Stellenvergabe funktional und *pragmatisch zielführend* ist. Der Vorteil der quantifizierenden und kalkulierenden Praktiken gegenüber zeitintensiven qualitativen Bewertungspraktiken liegt darin, dass sie sich der zahlenförmigen Repräsentation bedienen, die häufig auch in Grafiken visualisiert werden kann und somit die Fiktion der Vergleichbarkeit erzeugen. Die jeweiligen Scores, Zitationsindizes etc. sind schnell zu erfassen. Sie stellen einen

intersubjektiv nachvollziehbaren und gemeinsamen Bezugspunkt der Bewertung dar, der den Anschein einer verfahrenstechnischen Gleichbehandlung erweckt. Das erscheint fair oder gerecht, in dem Sinne, dass diese Kriterien als unempfindlich gegenüber dem Einfluss subjektiver Faktoren, wie beispielsweise Sympathien, Befangenheiten bzw. Bevor- oder Benachteiligungen erachtet werden. Der Nimbus der neutralen Urteilsposition scheint so gewahrt und die Legitimation durch Verfahren gestärkt. Indem die Metriken, oder konkreter die visualisierten Zahlen, Komplexität reduzieren, erscheinen Quantifizierungen für eine pragmatische Handhabung der gefragten Begutachtungen trotz ihrer genannten und kritisierten Limitationen unverzichtbar.

Dennoch bleibt die Funktionalität der Metriken mit Blick auf die Effektivität als Unterstützung des *Selektions- und Entscheidungsmechanismus* fraglich. Die Befragten äußern Zweifel, ob auf diese Weise richtig, also angemessen, bewertet wird.

“People are not assessed by a couple of numbers. And that one needs- to actually understand the value of someone, or of research proposal, one has to think about and actually know what the person has contributed, rather than just looking at the-you know, the titles of the journals in which this person has published” (Frau Simons, Prof., Abs. 2).

Zahlen sind nicht alles und es bleibt fraglich, ob sie die zu evaluierende Qualität aussagekräftig indizieren. An dieser Stelle zeigt sich, dass sich die befragten Lebenswissenschaftler/-innen der Grenzen der Quantifizierung und ihrer numerischen Repräsentation bewusst sind.

Wie gehen sie damit um und welche Maßnahmen fordern die Befragten, um eine dem Bewertungsgegenstand gerecht werdende und zugleich verfahrensförmig und damit legitime Aussagekraft zu erzeugen? Die Befragten fordern vor diesem Hintergrund, mehr und gründlicher zu lesen, die Indizes zu kontextualisieren und zusätzlich weitere Bewertungsparameter zu berücksichtigen.

“I think that the problem of the numbers is that they are absolutely dry, and they come with no context. And they come with essentially no context. So if you add a form of context that can be given by two or three experts of the field, maybe that starts to be eventually useful. But it cannot be used as a single way of classifying people. This is what bothers me” (Frau Gresillion, Prof., Abs. 88).

Im Kontext dieser von den Befragten geäußerten Forderungen lohnt sich ein Blick auf den Prozess der Personalrekrutierung, da hier die diversen Anforderungen wie Publikationen, Drittmittel, Lehre sowie persönliche und soziale Kompetenzen, Leitungs- und Teamfähigkeit zueinander ins Verhältnis gesetzt werden müssen. Zur Illustration der relevanten *sozialen Dimensionen* eignet sich folgende Interviewpassage:

“I think one important aspect is the human aspect. Whether for people who join my lab or more senior people joining the department at the PI level or so, an evaluation of the, you know, social aspect of that person is important, too, because it's important, we work together, and that interaction can be productive, that the work atmosphere is convivial. So I think, these sorts of things judged during interview on site are not negligible aspects of recruitment” (Frau Simons, Prof., Abs. 14).

Die Grenzen der Quantifizierung werden von den Befragten thematisiert und problematisiert. In der angeführten Passage geschieht dies mit Bezug auf „soziale Aspekte“, die sich einer Quantifizierung entziehen. Es kommen deshalb weiterhin qualitative Verfahrenselemente wie Bewerbungsvorträge und Bewerbungsgespräche zum Einsatz und diese werden auch als notwendig gerechtfertigt. Aller-

dings kommen auf dem Weg der Entscheidungsfindung dann im Dienst einer systematisch vergleichbaren Erfassung von quantitativen und qualitativen Aspekten Praktiken zur Anwendung, die man als Soziokalkulation bezeichnen kann.

Praktiken der Soziokalkulation

Die Aussagen der befragten Lebenswissenschaftler/-innen schildern auch Praktiken der Soziokalkulation, so dass die Reflexion und Kritik nicht auf die Frage des Ausmaßes der Quantifizierung in spezifischen Evaluationsmodi begrenzt werden kann. Daraus ergibt sich die Frage, inwiefern die Befragten eine Perspektive einnehmen, die über die kritische Betrachtung einzelner Verfahren hinausgeht?

Charakteristisch für die *Soziokalkulation*, im Unterschied zur gängigen Kalkulation, ist nach Vormbusch, dass die Trennung „zwischen messend-objektiven und kommunikativ-subjektiven Verfahren“ überwunden und beide Mechanismen miteinander verschränkt werden (Vormbusch 2012, S.206). Das Soziale wird messbar gemacht und das Gemessene wird somit für den spezifischen Kontext individuell validiert und mit einem verrechenbaren Sinn versehen. In den Bewertungsverfahren zur Personalrekrutierung manifestiert sich gerade diese Verbindung verschiedener Sozialtechniken, wenn als sogenannte ‚hard facts‘ Leistungsnachweise wie Publikationen, Patente, Spin-offs, Drittmittelakquise etc. gezählt, kalkuliert und nach Impact Faktoren, Größe und Menge gewichtet werden; und wenn parallel als soft skills persönliche, kommunikative und soziale Kompetenzen, wie Leitung, Lehre etc. im Gespräch eruiert und gegebenenfalls getestet werden und in der nachfolgenden Diskussion eingeschätzt, gelistet und schließlich gewichtet werden. Entsprechend der Praxis frühere wissenschaftliche Produktivität und Erfolge als Indizien für zukünftig zu erwartende Leistungen zu betrachten, bedient und erzeugt die Soziokalkulation hier systematisch eine „soziale Rationalität“ (Vormbusch 2012, S.206), in der die *Erfassung und Messung vergangener Leistungen mit einer Diagnose und Prognose* für zukünftige Leistungen verknüpft sind. Dabei werden die subjektiv-kommunikativen Prozesse vorstrukturiert und kategorisiert, um systematisch abstrahiert in die Kalkulation eingespeist werden zu können. In diesem Sinne ist die Soziokalkulation nicht nur *erhebend und erfassend, sondern auch konstruktiv* wirksam, indem sie Erwartungshorizonte modelliert.

Derart tragen die soziokalkulativen Praktiken der Kritik der Befragten gewissermaßen Rechnung, denn sie leisten eine Vervielfältigung und Kontextualisierung, indem sie individuelle, kontextbezogene und soziale Bewertungsparameter, zum Beispiel mit Blick auf die Teampassung, mit aufnehmen. Dabei werden die subjektiv-kommunikativen Prozesse vorstrukturiert und systematisch abstrahiert. Die „Fähigkeit kalkulativer Praktiken, soziale Phänomene in organisierbare, komplexitätsreduzierte und handhabbare Größen zu überführen“ (Vormbusch 2012, S.223) kommt zum Tragen. Die kalkulative Erfassung, der zunächst qualitativ eruierten sozialen Aspekte, erfordert und bedingt somit deren Kategorisierung und Standardisierung. Dadurch werden auch die sozialen Aspekte anschlussfähig für etwaige Skalierungen und Quantifizierungen beispielsweise für akkumulierte soft skill scores oder grafische Darstellungen, um in weitere Kalkulationen eingespeist werden zu können. Dabei bedienen sie auch normative Anforderungen an faire Bewertungsverfahren; denn sie erlauben eine verfahrensmäßige Gleichbehandlung und erzeugen somit die Fiktion der Kommensurabilität (Vormbusch 2012, S.208).

Hier wird erkennbar, dass es nicht des Glaubens an eine scheinbare Objektivität und Neutralität der Zahlen bedarf – wie die Kritik der Befragten illustriert – für dessen Wirkmächtigkeit reicht es, dass sie als „pragmatisches Handlungsprogramm“ (Vormbusch 2012, S.227) funktionieren. Als solches kann

auch die Soziokalkulation Antrieb der Eigendynamik weiterer quantifizierender und standardisierender Elemente im wissenschaftlichen Evaluationssystem sein.

Angesichts der Kritik der Befragten an einer Quantifizierung und deren Funktionalität für eine Bewertung wissenschaftlicher Leistung stellt sich die Frage, ob die Soziokalkulation selbst als Teil der Quantifizierung zu bewerten ist; oder ob sie genau an den Grenzen der Quantifizierung – soweit sie auf numerischen Operationen basiert – ansetzt und diese mit Hilfe von begrifflicher Kategorisierung und Standardisierung zu überwinden sucht? Folglich geht die Frage nach den Grenzen der Quantifizierung mit der Frage nach den Grenzen der kategorisierenden, standardisierenden und soziokalkulativen Erfassung einher (Heintz 2010, Krüger, Reinhart 2017). Quantifizierung ist in ihrer Potenz, gesellschaftliche Wirklichkeit zu erfassen und zu beschreiben, insofern begrenzt, als sie notwendig mit der Reduktion somit einem Informationsverlust und der Fragmentierung im Zuge numerischer Repräsentationen einhergeht.

Das Spannungsverhältnis zwischen Exzellenz und Mittelklasse

Mit Blick auf die gesellschaftliche Relevanz jener Bewertungsprozeduren für wissenschaftspolitische Governance gilt es aus soziologischer Perspektive die Limitationen der Kategorisierung, Standardisierung und der soziokalkulativen Praktiken zu hinterfragen. Daran, welche Dimensionen und Kriterien der Personalrekrutierung kritisch hinterfragt und erörtert werden, lässt sich ablesen, welche der für die Personalauswahl relevanten Aspekte weder durch die quantifizierte Evaluation von Publikationen und Drittmittelinwerbung, noch durch die Soziokalkulation angemessen erfasst werden. Eine Befragte nennt und gewichtet die für die Rekrutierung von Assistant-Professor/-innen zentralen Kriterien folgendermaßen: 50–60% wissenschaftliche Qualifikation, 20% Führungskompetenz, 20% Lehrkompetenz, 10% Großmut und Konzilianz sich institutionell einzubringen.

„[S]o, this should be at least 50 percent. Maybe up to 60 percent. Because I think that we need to have someone who has some leadership. Because I saw some people, they are not bad scientists, they are reasonable scientists, but they have absolutely no leadership. [A]ll I would say that at least half of the problems that we have with researchers are manager-problems. And the other thing that should be taken into account, is that [...] it has to be someone a little bit generous. If you are not generous in science, okay, it's good for the person, I think it's- I see some example on ?XX?, they are extremely productive, they are good, they are- but they don't give a damn for the institution. So there has to be a kind of generosity in their way of dealing with the institution. And then the last thing is that this person needs to be interested in teaching. At least a minimum of pedagogical qualities“ (Frau Gresillion, Prof., Abs 110–112).

Dabei offenbart sich anhand der von den befragten Wissenschaftler/-innen dargelegten Bewertungs- und Leistungsvermessungsprozesse zusätzlich zu den Vor- und Nachteilen der Quantifizierung und Formalisierung, in welchem Maße die Evaluationsverfahren in der Wissenschaft selber als wissenschaftspolitische Steuerungsinstrumente Qualitäts- und Legitimationsanforderungen unterworfen sind. Auch hier bietet die Kritik der Befragten eine Grundlage für die kritische Prüfung der Funktionalität jener Bewertungsmechanismen.

Mehrere Befragte kritisieren eine einseitige Fokussierung der Förderung auf Exzellenz:

“Essentially, these committees (Berufungskommissionen) are looking for what they would define with the word ‘excellence’. Again, that does not necessarily produce a match but usually, people who are very outspoken and engaging and presenting well with various small things, they are doing better on these things” (Frau Todorova, Prof, Abs. 107).

Auf die Frage, welche Qualitätskriterien für die Bewilligung von Projektanträgen entscheidend sein sollten, formuliert ein Befragter:

„Das müsste sein die Innovation, was bringt das? Welche technischen, gesellschaftlichen Fragestellungen und Probleme löst man damit?“ (Herr von Stade, FH, Abs. 13).

Damit fordert er gewissermaßen einen alternativen Deutungshorizont von exzellenter Qualität ein, die er an der relativen Kapazität bzw. dem Potenzial festmacht, zur Lösung relevanter Probleme beizutragen.

Vergleichbare kritische Reflexionen sind auch mit der Forderung nach einer stärkeren und kontinuierlichen Breitenförderung der Forschung verknüpft:

“So the point is more, okay, should science be made only by a few original guys, and in that case, okay, we will make a lot of economy, but I am not sure that we will go very far. Or should we really be much more encouraging for this kind of middle class of scientists who are really good. So, for these people, how can we encourage creativity and originality? And maybe this is for those people [middle-class scientists], that it's important not to have this kind of issue, you have to have a Cell paper, otherwise you are- because it encourages a lot of- yes, there is a lot of backdrop to this kind of approach and of categorising people only with respect to their publication” (Frau Gresillion, Abs. 54).²

Die Befragte stellt in Frage, ob die Fokussierung auf die Förderung von Exzellenz langfristig für die Wissenschaft ertragreich ist. Der Förderung einer exzellenten aber notwendig eng begrenzten Elite stellt sie die sogenannten Mittelklasseforscher/-innen gegenüber, die solide Arbeit machen und deren Originalität und Kreativität die Befragte für ebenso förderungswürdig erachtet. Sie betont darüber hinaus, dass eine wissenschaftspolitische Fokussierung auf Exzellenz einseitige Anreize setzt, die auf Kosten der „Mittelklasseforschung“ gehen können. In der Kritik der Befragten deutet sich eine spezifisch-reflexive Perspektive an, welche die Wissenschaft als einen konkreten gesellschaftlichen Teilbereich umfasst und dessen Funktionalität fokussiert. In dieser Perspektive kommt der sogenannten „Mittelklasse“-Forschung ebenso sehr wie der dazu in Abgrenzung gesetzten exzellenten, herausragenden Forschung eine tragende Rolle als Motor der Wissensproduktion und Quelle für Erkenntniszuwachs und technologische Weiterentwicklung zu, die sodann der Gesamtgesellschaft und ihrer Volkswirtschaft kontinuierlich zu Gute kommen und deren Zukunftsfähigkeit mit gewährleisten kann. Aus dieser Perspektive scheint es naheliegend, dass Exzellenz nicht als alleiniger Entscheidungsmaßstab dafür herangezogen werden kann, welche Forschung grundsätzlich förderungswürdig erscheint. Sofern die Aufgabe der Wissenschaft und Forschung darin besteht, dauerhaft und langfristig gesellschaftliche Entwicklungspotentiale bereitzustellen, bedarf es ebenso der Lehre, der Ausbildung, der institutionellen Einbindung und organisierten Förderung einer Breitenforschung bzw. „Mittelklassen“-

² Die Befragte war mehrere Jahre Dekanin und positioniert sich in ihren weiteren Erläuterungen folgendermaßen zur Mittelklasse: „I will call them middle class, okay? And I am putting myself in this middle class, scientists“ (Abs. 114).

Wissenschaft, welche den stetigen Forschungsfundus erarbeitet und nutzt, und aus dem dann gegebenenfalls Außergewöhnliches und Bahnbrechendes erwachsen und entwickelt werden kann.

„Also ein Riesenwunsch wäre natürlich, dass man einen Mittelbau hätte, der grundfinanziert ist, dass die Leute einfach länger hierbleiben können und Mitarbeiter dann Erfahrung haben“ (Herr von Stade, Abs. 49).

Die Befragten mobilisieren eine als allgemeine Kritik verstandene Unterscheidung zwischen einer kleinen hochspezialisierten, die Spitzenmarke setzenden wissenschaftlichen „Exzellenz-Elite“ und den zahlreichen wissenschaftlich ausgebildeten und tätigen „Wissensarbeitern“, die den gesamtgesellschaftlichen Bedarf an wissenschaftlicher Arbeit abdecken und eine wissenschaftliche Mittelklasse repräsentieren (vgl. auch Elias et al. 1982; Whitley 2000; Whitley et al. 2010)³.

Das Rechtfertigungsimperativ einer gesellschaftsorientierten Breitenforschung

Die von den wissenschaftlichen Akteuren formulierte Kritik lässt sich mit Boltanski und Thévenot (2007), wie im Weiteren gezeigt werden soll, als eine Mobilisierung des Staatsbürgerlichen interpretieren. Den Autoren zufolge sind die Akteure grundsätzlich mit kritischen Kapazitäten ausgestattet, die es ihnen ermöglichen, in Situationen des Disputs spezifische Rechtfertigungsmuster aufzurufen, die so dann als legitime Form von Kritik fungieren. In unserem Fall zeigt sich, dass die Akteure sich kritisch mit Quantifizierung auseinandersetzen, sie hinterfragen, sie aber zugleich in pragmatischer Wertschätzung der Funktionalität quantifizierender Bewertungsmodi in einem variablen Maß praktizieren und damit gleichsam reproduzieren. Dieses ambivalente Verhältnis äußert sich auf praktischer Ebene zum einen in Formen der Soziokalkulation, welche unter anderem die Handlungsfähigkeit der Akteure im Alltag sicherstellt. Darüber hinaus mündet sie in Formen der konkreten Wissenschafts- und Gesellschaftskritik, welche die Kluft zwischen Exzellenz und Mittelklasse offenbart. Um dieses scheinbar ambivalente Verhältnis zwischen einem pragmatischem Vorgehen der Akteure einerseits (Partizipation und Reproduktion quantifizierender Handlungspraxis) und Gesellschaftskritik andererseits besser verstehen zu können, bedarf es einer genauer differenzierenden Bestimmung der Rechtfertigungsmuster, auf welche die Akteure in ihrer Kritik zurückgreifen. In der von den befragten Wissenschaftler/-innen formulierten Kritik wird die gesellschaftliche Dimension und wissenschaftspolitische Stoßrichtung deutlich, und zwar insofern sie spezifische Begründungsmuster und Rechtfertigungsordnungen aufruft. Über Gründe der Sicherung und Förderung hoher wissenschaftlicher Qualität und manifester Forschungserfolge hinaus werden von den Befragten Argumente vorgebracht, die eher auf eine kontinuierliche solide und zeitgemäße Versorgung der Gesellschaft mit stetig aktualisierter Wissensproduktion durch öffentlich finanzierte Forschung abzielen, als auf eine Reproduktion von wissenschaftlicher Exzellenz und damit einhergehend einer Exzellenz-Elite. Darüber hinaus werden mit Blick auf die Argumentation, eine andauernde Wissensproduktion zu gewährleisten, Forderungen nach einer Art Verteilungsgerechtigkeit innerhalb der verschiedenen Stoßrichtungen wissenschaftlicher Forschung durch die Befragten hervorgebracht. Durch die analytische Brille einer Soziologie der Kritik betrachtet, werden in den zuvor dargelegten Bewertungsverfahren, die als „Prüfungen“ fungieren, wissenschaftliche Gründe in der Kritik der Befragten, die zum Zweck der Validierung und Qualitätssi-

³ Auch diese Unterscheidung ist nicht unbekannt, sondern im professionalisierten Personalmanagement gängig, bei dem die Soziokalkulation systematisch zum Einsatz kommt, unter Schlagwörtern wie „Work horses“ (Leistungsträger) denen „Stars“ (Leuchttürme) gegenüberstehen sowie Gewissenhafte, die zuverlässig und genau abarbeiten, denen tendenziell unberechenbare, aber radikal-kreative Köpfe gegenüberstehen (vgl. Vormbusch 2012, S.206-218).

cherung angeführt werden, in eine Äquivalenzbeziehung zum Rechtfertigungsmuster des staatsbürgerlichen Gemeinwesens gebracht. Diese Rechtfertigung wird gegen eine Wissenschaftspolitik in Stellung gebracht, die scheinbar darauf setzt, dass sich im Strömungskanal der Spitzenförderung der notwendige Rekrutierungspool für die Spitze im Sinne einer „Mittelklasse“ von selber herausbildet. Diese Stoßrichtung der Kritik verharret somit nicht in dem Konflikt unterschiedlicher Rechtfertigungsordnungen. Indem die Forscher/-innen die wissenschaftspolitische Programmatik, insbesondere das aktuelle Paradigma der Exzellenzförderung kritisieren, hinterfragen sie diese auf jenem Terrain, das ansonsten von Akteuren aus der Wissenschaftspolitik eingenommen und beansprucht wird. Durch die Bezugnahme auf das Rechtfertigungsmuster des staatsbürgerlichen Gemeinwesens eröffnen sie die Möglichkeit, vom Konflikt in den Streit überzugehen.⁴ Dabei verwenden sie Argumentationsformen, die aus der Governancelogik stammen, welche auf die Ineffizienz und Dysfunktionalität der Personalrekrutierung hinweisen und als Konsequenz problematische Effekte in der sozialen Stratifikation der wissenschaftlichen Community erzeugt. Solche Konsequenzen werden beispielsweise dann erzeugt, wenn unterhalb der exzellenten Spitze aus Großteils öffentlichen Geldern finanziertes, langjährig berufserfahrenes, hochqualifiziertes und spezialisiertes Personal aus dem Wissenschaftssystem herausfällt, und es gleichzeitig an geeigneten Übergangsmöglichkeiten in andere gesellschaftliche Teilbereiche mangelt.

„[E]s ist immer ein bisschen so, ich habe keine Probleme auf etwas Kompetitives hinzuarbeiten, wenn ich das will. Ich tue mein Bestes, aber wenn dann, wenn man dann scheitert und es ist keine Lösung mehr da, man ist völlig, es gibt keine Ausbruchsmöglichkeiten mehr. [...] Also, ich sage einfach nur, dass diese Systeme nicht mehr ineinander abgestimmt, aus meiner Sicht, weil immer mehr erwartet wird. Die Zeiträume werden immer länger, die PhDs werden immer länger, die Postdocs werden immer länger und komplexer und wenn die dann zurückkommen, dann haben sie genau ein window von zwei Jahren, wenn es da nicht klappt, sind sie tot (Herr Mühlhausen, Prof., Abs. 53).

Die von den Befragten formulierte Kritik kann mit Blick auf die hervorgebrachten Rechtfertigungsmuster folgendermaßen interpretiert werden: Es wird gefordert, dass die Gestaltung und die Gewichtung der verschiedenen Bewertungselemente und Prozesse, die in Bewertungsverfahren zur Anwendung kommen, auf wissenschaftliche Güte zielen und dabei die Balance zu Aspekten des Staatsbürgerlichen wahren. Einer derartigen Interpretation zufolge kann die Aushandlung seitens der wissenschaftlichen Akteure, welches Äquivalenzprinzip in den Bewertungsprozessen der Wissenschaft geltend gemacht werden soll, als Bedingung dafür verstanden werden, dass diese Evaluationsprozeduren als Prüfungsformate den Qualitäts-, Kommensurabilitäts- und Legitimationserfordernissen gerecht werden müssen, die sowohl in der wissenschaftlichen Community als auch in einer demokratischen Gesellschaft an sie gestellt werden; denn zumindest die öffentlich finanzierte Forschung verspricht auch im Dienste des Gemeinwohls ein valides Wissen zu erarbeiten und zur gesellschaftlichen Nutzung bereitzustellen.

⁴ Vergleiche Boltanski und Thévenot (2007, S.54 ff.): Streit, zwei Parteien beziehen sich auf dieselbe Rechtfertigungsordnung; Konflikt, zwei Parteien beziehen sich auf unterschiedliche Rechtfertigungsordnungen. Die Prüfung klärt, sodann welche Ordnung als Äquivalenzprinzip geltend gemacht wird.

Fazit

Die vorliegende Studie hat sich dezidiert mit Quantifizierungsdynamiken in Bewertungs- und Peer-Review-Verfahren in den Lebenswissenschaften auseinandergesetzt. Die Ergebnisse der empirischen Studie zeigen, dass eine als Quantifizierung verstandene Entwicklung im Rahmen wissenschaftlicher Evaluationen von den befragten Wissenschaftler/-innen zwar ausgiebig bemängelt und kritisch betrachtet, aber nicht per se als problematisch erachtet wird; denn die Quantifizierung erfordernden Praktiken des (Ver-)messens sind durchaus als wissenschaftsadäquate Vorgehensweise anerkannt. Eine vorschnelle Beurteilung der Kritik der Befragten als ein Imperativ gegen Quantifizierung greift daher zu kurz und unterschätzt zugleich die Fähigkeit der Forschenden zur Formulierung einer „Systemkritik“, die weit über die eigene Position als Forschende/r hinausgeht. Die Feinanalyse der Äußerungen der Befragten zeigt, dass an Bewertungsverfahren in der Wissenschaft auch ‚politische‘ und gemeinwohlorientierte Mindeststandards mit Blick auf Partizipationsmöglichkeiten und Fragen der Verteilungsgerechtigkeit angelegt werden. Letztere stehen einer zu starken Binnen-Stratifizierung von Position und Status innerhalb der wissenschaftlichen Community entgegen, die in der „Exzellenz-Elite“ ihren Ausdruck findet, welche die Breite und Tiefe der Forschungsmöglichkeiten aller Forscher/-innen mitbedingt und die Ausstattung der „Mitteklasseforschung“ determiniert. In dieser Kritik offenbart sich somit eine gesellschaftspolitische Ebene der Problematisierung von Quantifizierung, bei der die Akteure auf eine staatsbürgerliche Polis rekurrieren, die sogleich als legitimes Rechtfertigungsmuster fungiert.

Literatur

- Angermüller, Johannes. 2011. Wissenschaft zählen. Regieren im digitalen Panopticon. In *Sichtbarkeitsregime: Überwachung, Sicherheit und Privatheit im 21. Jahrhundert*, Hrsg. Leon Hempel, Susanne Krasmann und Ulrich Bröckling, Leviathan Sonderheft, 174–190. Wiesbaden: VS Verlag.
- Boltanski, Luc und Laurent Thévenot. 2007. *Über die Rechtfertigung. Eine Soziologie der kritischen Urteilskraft*. Hamburg: Hamburger Edition.
- Elias, Norbert, Herminio Martins, und Richard Whitley. 1982. (Hrsg.) *Scientific establishments and hierarchies*. Dordrecht: D. Reidel Publishing Company.
- Fourcade, Marion und Kieran Healy. 2013. Classification situations: Life-chances in the neoliberal era. *Accounting, Organizations and Society* 38:559–572.
- Heintz, Bettina. 2010. Numerische Differenz. Überlegungen zu einer Soziologie des (quantitativen) Vergleichs. *Zeitschrift für Soziologie* 39:162–181.
- Hendriks, Barbara, Martin Reinhart und Cornelia Schendzielorz. 2018. Förderung und Evaluation der Forschung in den Lebenswissenschaften in der Schweiz – eine Interviewstudie. *The growth of science: Auswirkungen für die Forschungsevaluation und -förderung in der Schweiz. Politische Analyse und Empfehlungen des Schweizerischen Wissenschaftsrates SWR. Empirische Untersuchung von B. Hendriks, M. Reinhart und C. Schendzielorz*, 55–99. Bern: Swiss Science Council.
- Hornbostel, Stefan. 2015. Das Zitat als Währung. In *Zitat, Paraphrase, Plagiat. Wissenschaft zwischen guter Praxis und Fehlverhalten*, Hrsg. Christiane Lahusen und Christoph Marksches, 307–316. Frankfurt am Main/ New York: Campus.
- Kraft, Alison. 2013. New Light Through an Old Window?: The „Translational Turn“ in Biomedical Research: A Historical Perspective. In *Translational Medicine. The Future of Therapy?* Hrsg. James Mittra und Christopher-Paul Milne, 19–55. Singapore: Pan Stanford Publishing.

- Krüger, Anne K. und Martin Reinhart. 2016. Wert, Werte und (Be)Wertungen. Eine erste begriffs- und prozesstheoretische Sondierung der aktuellen Soziologie der Bewertung. *Berliner Journal für Soziologie* 26:485–500.
- Krüger, Anne K. und Martin Reinhart. 2017. Theories of Valuation – Building Blocks for Conceptualizing Valuation Between Practice and Structure. *Historical Social Research* 42:263–285.
- Lamont, Michèle. 2012. Toward a Comparative Sociology of Valuation and Evaluation. *Annual Review of Sociology* 38:201–221.
- Macleod et al. 2014. Biomedical research: increasing value, reducing waste. *The Lancet* 383:101–104.
- Reinhart, Martin. 2012. *Soziologie und Epistemologie des Peer Review*. Baden-Baden: Nomos.
- Vormbusch, Uwe. 2012. *Die Herrschaft der Zahlen. Zur Kalkulation des Sozialen in der kapitalistischen Moderne*. Frankfurt am Main: Campus.
- Whitley, Richard. 2000. *The Intellectual and Social Organization of the Sciences*. Oxford: Oxford University Press.
- Whitley, Richard, Jochen Gläser, und Lars Engwall. 2010 (Hrsg.). *Reconfiguring Knowledge Production: Changing authority relationships in the sciences and their consequences for intellectual innovation*. Oxford: Oxford University Press.
- Wilholt, Torsten. 2012. *Die Freiheit der Forschung. Begründungen und Begrenzungen*. Berlin: Suhrkamp.