

Editorial 2019

Gunnar Otte:

Ein Blick in die Black Box des Review-Verfahrens

<https://doi.org/10.1515/zfsoz-2019-0001>

„Die *Zeitschrift für Soziologie* ist eine Zeitschrift für das gesamte Fach. Sie veröffentlicht Beiträge aus allen Bereichen der Soziologie. Dies bezieht sich sowohl auf die Vielfalt der Forschungsgebiete als auch auf die Pluralität von Schulen in der empirischen Forschung und der soziologischen Theoriebildung. Die Gründungsherausgeber von 1971 konzipierten die *ZfS* als ein repräsentatives Organ ihres Faches, das quer zu den vorhandenen theoretischen Orientierungen und empirischen Forschungsgebieten allen Soziologinnen und Soziologen offen stehen soll.“ So lautet das Selbstverständnis der *Zeitschrift für Soziologie* auf ihrer Homepage. Institutionell versucht die *ZfS* diese Zielsetzung durch zwei Verfahren zu erreichen: durch ein rotierendes, pluralistisch besetztes Gremium von derzeit fünf Herausgebern und Herausgeberinnen; und durch zwei Fachgutachten, die pro Aufsatz eingeholt werden. In beiden Fällen findet die Begutachtung anonym im „double blind“-Verfahren statt. Die Autoren und Autorinnen erhalten in diesem aufwändigen und mitunter als vorbildlich erachteten Verfahren der Qualitätssicherung bis zu sieben Beurteilungen ihres Manuskriptes rückgemeldet.

Doch welche Bewertungs- und Selektionsprozesse finden innerhalb dieses Systems statt? Kann die *ZfS* als „repräsentatives Organ“ der (deutschsprachigen) Soziologie angesehen werden? Im strikten Sinne würde dies voraussetzen, dass Aufsätze aus allen Teilgebieten der Soziologie proportional zu ihrem Personalvolumen bei der *ZfS* eingereicht werden, dass die wissenschaftliche Qualität von Manuskripten in allen Teilgebieten identisch verteilt ist und dass im Begutachtungsprozess kein systematischer Bias jenseits anerkannter Qualitätskriterien auftritt. Wollte man dies empirisch untersuchen, müsste man die Grundgesamtheit soziologischer Forschungstätigkeit quantifiziert werden – ein aufwändiges Unterfangen. Zum anderen wäre zu objektivieren, inwieweit wissenschaftliche Qualität über die Forschungsfelder der Soziologie ungleich verteilt ist – ein nicht minder schwieriges und Streitbares Unterfangen.

Derart weitreichend ist der Anspruch dieses Editorials nicht. Doch soll es den Prozess der Auswahl von Manu-

skripten aus einer in der Peer Review-Forschung selten anzutreffenden Binnenperspektive transparenter machen (vgl. Hirschauer 2004). Dies möchte ich am Ende meiner vierjährigen Amtszeit als Mitherausgeber der *ZfS* auf der Basis statistischer Daten und subjektiver Wahrnehmungen tun. Das Interesse an dieser Frage ist nicht neu. Bereits Sahner (1982) ging der Selektivität des Peer Reviews anhand eines Vergleichs aller eingegangenen und schließlich publizierten Manuskripte in der *ZfS* der Jahrgänge 1972 bis 1980 nach. Sortiert nach theoretischen Ansätzen zeigte sich etwa, dass Beiträge der kritischen Theorie, des historischen Materialismus und der verstehenden Soziologie im Review-Prozess überproportional scheiterten, während Manuskripte funktionalistischer Provenienz in hohem Maße abgedruckt wurden. Nicht lösen konnte Sahner das Rätsel, ob dafür Qualitätsunterschiede oder andere Selektivitäten der Begutachtung verantwortlich waren. Er zitiert jedoch das Editorial in Heft 1 des Jahrgangs 1975, in dem die Herausgeber vielen Einreichungen marxistischer Prägung „eindeutig mindere Qualität“ attestieren.

Für alle Manuskriptenreichungen der Jahre 2015 und 2016 habe ich ebenfalls statistisch erfasst, wie sie beurteilt und prozessiert wurden. Insgesamt handelt es sich um 157 Fälle. Nicht enthalten sind darin die Beiträge der Vorjahre, die nach einer Überarbeitung erneut eingereicht wurden, sowie zwei Manuskripte, die aufgrund mangelnder Eignung nicht in den Review-Prozess gelangten („desk rejection“). Während Sahner die Manuskripte u. a. nach Forschungsfeldern und theoretischen Ansätzen kodierte, beschränke ich mich auf eine simple Einteilung vier typischer Textgattungen. Theoriebeiträge sind Texte, in denen keine systematisch erhobenen Daten ausgewertet werden. Dabei handelt es sich teils um Beiträge zur allgemeinen soziologischen Theoriebildung, teils um Arbeiten spezieller Soziologien, z. B. zur Organisationssoziologie. Unter den empirischen Beiträgen werden diejenigen als quantitativ klassifiziert, die statistische Datenanalysen mit großen Fallzahlen vornehmen. Typischerweise handelt es sich um Sekundärdatenanalysen, z. B. des Sozioökonomischen Panels. Als qualitative Forschung gelten Beiträge, in denen die interpretative Auswertung von Text- oder Beobachtungsdaten im Zentrum steht. Aufsätze, die Datenarten oder Auswertungslogiken beider Forschungsstile kombinieren (z. B. Inhaltsanalyse, QCA/FSA), werden als

*Korrespondenzautor: Gunnar Otte, Johannes Gutenberg-Universität Mainz, Institut für Soziologie, Jakob-Welder-Weg 12, 55128 Mainz, E-Mail: gunnar.otte@uni-mainz.de

Methodenkombination bezeichnet. Nicht separat erfasst wird reine Methodenforschung; sie wird entweder der quantitativen oder der qualitativen Kategorie zugeordnet.

Tab. 1: Ersteingereichte und begutachtete Manuskripte nach Gattungen 2015/16

Manuskriptgattung	n	%
Theorie	48	30,6
Empirie: Qualitativ	33	21,0
Empirie: Quantitativ	69	43,9
Empirie: Methodenkombination	7	4,5
Gesamt	157	100,0

Wie Tabelle 1 zeigt, setzten sich die Ersteinreichungen zu rund 70 % aus empirischen und zu 30 % aus theoretischen Beiträgen zusammen. Diesem Verhältnis liegt ein allgemeiner Trend in Richtung empirischer Beiträge in großen allgemein-soziologischen Zeitschriften wie der *ZfS* und *KZfSS* zugrunde (vgl. Kopp et al. 2012: 303). Unter den datengestützten Einreichungen sind quantitative Beiträge doppelt so stark vertreten wie qualitative. Methodenkombinationen sind rar.

Von den 157 Einreichungen schafften es 49 Beiträge – nach einer mehr oder minder großen Überarbeitung – zur Publikation in der *ZfS*. Dies entspricht einer Quote von 31 %. In Tabelle 2 wird jedoch erkennbar, dass die Annahmequote nach Manuskriptgattung variiert: Sie liegt für quantitative Beiträge bei 42 %, für Theoriebeiträge bei 27 % und für qualitative Beiträge bei 15 %. Der Zusammenhang ist mittelstark ($V=0,23$) und statistisch signifikant ($\chi^2=8,129$; $p<0,05$). In den publizierten Ausgaben der *ZfS* machen empirische Beiträge somit fast 75 % aus – eine Fortschreibung des Trends, den Kopp et al. (2012: 303) für die Zeit von 1972 bis 2010 nachweisen. Und es dominiert die quantitative Forschung in den publizierten Heften noch stärker (59 %) als im Manuskripteingang (44 %). Aufgrund der kleinen Fallzahlen sollten die Werte aber nicht überinterpretiert werden; bereits kleine Unterschiede in den Entscheidungen können das Bild stark verändern.

Tab. 2: Publierte Manuskripte nach Gattungen

Manuskriptgattung	n	Annahmequote in %
Theorie	13	27,1
Empirie: Qualitativ	5	15,2
Empirie: Quantitativ	29	42,0
Empirie: Methodenkombination	2	28,6
Gesamt	49	31,2

Fast alle publizierten Beiträge wurden erst nach einer Überarbeitung angenommen. Über die Hälfte (55 %) der Publikationen war zuvor zu einer gründlichen Revision und Neueinreichung aufgefordert worden (vgl. Tabelle 3). Bei 39 % der publizierten Beiträge erfolgte eine kleinere Überarbeitung. Lediglich im Theoriebereich waren drei Manuskripte so überzeugend, dass sie direkt in die Publikation gingen. Die Variation der Überzeugungskraft der Manuskripte ist in der soziologischen Theorie somit besonders groß.

Tab. 3: Publierte Manuskripte nach Überarbeitungsstatus

Manuskriptgattung	Ohne Revision	Kleine Revision	Große Revision	Gesamt
Theorie	23,1 % (3)	38,5 % (5)	38,5 % (5)	100,1 % (13)
Empirie: Qualitativ	0,0 % (0)	40,0 % (2)	60,0 % (3)	100,0 % (5)
Empirie: Quantitativ	0,0 % (0)	41,4 % (12)	58,6 % (17)	100,0 % (29)
Empirie: Methodenkombination	0,0 % (0)	0,0 % (0)	100,0 % (2)	100,0 % (2)
Gesamt	6,1 % (3)	38,8 % (19)	55,1 % (27)	100,0 % (49)

Welches Bewertungsgeschehen liegt dem in diesen empirischen Verteilungen resultierenden Entscheidungsprozess zugrunde? Den zwei Gutachtenden liegt ein vierstufiges Klassifikationsschema vor:

- Das Manuskript ist publikationswürdig und kann vorbehaltlich geringfügiger Modifikationen sofort veröffentlicht werden.
- Das Manuskript ist publikationswürdig, bedarf allerdings vor der Veröffentlichung noch der Überarbeitung einiger Schwächen.
- Das Manuskript ist potenziell publikationswürdig, kann in der vorliegenden Form aber nicht zur Publikation akzeptiert werden. Die Autoren sollten zu einer umfänglichen Überarbeitung und Neueinreichung des Beitrags ermutigt werden. Eine erfolgreiche Überarbeitung ist den Autoren zuzutrauen.
- Das Manuskript ist nicht publikationsfähig und vermutlich auch durch umfangreiche Überarbeitungen nicht zu retten.

Entscheidend für den *ZfS*-internen Diskussionsprozess ist neben diesem Rating die ausführliche textliche Begutachtung, die sich oft über mehrere Seiten erstreckt. Die individuellen Voten der fünf Herausgebenden orientieren sich an einem etwas differenzierteren Bewertungsschema, das

aber im Prinzip auf die vier dargestellten Stufen reduziert werden kann. Darüber hinaus enthalten die Voten eine schriftliche Begründung variabler Länge. Auf der Basis der Gutachten und der Voten trifft das Herausbergremium im mündlichen Diskurs eine einvernehmliche Entscheidung über Annahme, Ablehnung bzw. Ausmaß und Ausrichtung der empfohlenen Überarbeitung.

Wie groß ist der Konsens in den Gutachten und Voten? Um dies zu analysieren, werden die vier Bewertungsstufen des obigen Schemas mit einer Metrik von 1 (Ablehnung) bis 4 (vorbehaltlose Annahme) quantifiziert. Sämtliche textuellen Bewertungen werden dabei ignoriert. Der Analyse liegt also eine vereinfachte und ggf. verzerrte Darstellung des faktischen Entscheidungsprozesses zugrunde. Für die Stichprobe der 157 Manuskripte liegt das mittlere Urteil der Gutachtenden bei 2,15 (auf der Skala von 1 bis 4). Die individuellen Voten der Herausgebenden fallen etwas negativer aus und liegen im Schnitt bei 1,65. Die Letztentscheidung über den Umgang mit dem Manuskript nach der gemeinsamen Aussprache liegt leicht darunter (1,54). Die Diskrepanz zwischen den Gutachtenden und Herausgebenden lässt sich teilweise dadurch erklären, dass bei der Herausgabe stärker mitbedacht wird, ob ein Manuskript thematisch zu spezialisiert für eine allgemein-soziologische Zeitschrift wie die *ZfS* sein könnte. Manche grundsätzlich guten Beiträge werden aus diesen Gründen abgelehnt und zur Einreichung bei einem spezialisierten Journal empfohlen.

Auffälliger sind die Unterschiede im Beurteilungskonsens. Die Korrelation der Urteile der zwei Gutachtenden eines Manuskriptes beträgt $r=0,24$ ($p<0,01$). Die Gutachten sind weit von einem Konsens entfernt, weisen aber tendenziell in dieselbe Richtung. Dieser Wert mag niedrig erscheinen, wenn man erwartet, dass die Reviewer im Sinne einer hohen Inter-Rater Reliabilität zu derselben Einstufung kommen sollten. In der Peer Review-Forschung sind derartige Werte aber nicht ungewöhnlich. Für acht Jahrgänge der *Zeitschrift für Internationale Beziehungen* wird eine Intra-class correlation (ICC) von 0,14 ermittelt (Weller 2004: 365). Eine Metaanalyse von 48 internationalen Studien berichtet eine durchschnittliche ICC von 0,34 (Bornmann et al. 2010). Die ICC liegt im Fall der *ZfS* mit 0,24 etwas darunter. Korrelationen dieser Größenordnung kommen zustande, weil die Gutachtenden von der Redaktion zum Teil komplementär ausgewählt werden, so dass ein Manuskript etwa von einer Methodenspezialistin und einem Gegenstandsexperten beurteilt wird. Hinzu kommt aber, dass die Peers von sich aus unterschiedliche Aspekte eines Beitrags ins Visier nehmen und zu abweichenden Einschätzungen kommen können. Wenn Reliabilitätskoeffizienten keinen Konsens anzeigen, ist dies für

die Qualitätssicherung insofern nicht unbedingt kritisch zu werten, als die Qualität des Manuskriptes mit breiterer Perspektivität bewertet wird (Hirschauer 2004).

Deutlich größer ist der Konsens zwischen den Voten der fünf Herausgebenden, die im Vorfeld der Sitzungen unabhängig voneinander erstellt werden. Die mittlere paarweise Korrelation ihrer Urteile liegt bei $r=0,52$ ($ICC=0,50$). Entsprechend selten gab es in den vergangenen Jahren große Dispute bei den Herausbertreffen. Die Diskussionen kreisten überwiegend um die Frage, wie einzelne Manuskripte verbessert werden können. Setzt man das durchschnittliche Urteil der beiden Fachgutachten mit der finalen Entscheidung am Ende eines Herausbertreffens in Beziehung, ergibt sich ein Zusammenhang von $r=0,45$. Substanziell bedeutet dies, dass die Stellungnahmen der Gutachtenden die Entscheidungsfindung der Herausbergerschaft stark anleiten, sie aber nicht determinieren. Dies entspricht durchaus dem Selbstverständnis der *ZfS* und ihrem Mechanismus der doppelten Qualitätssicherung, der externe und interne Urteile miteinander verbindet.

Kommen wir zurück zur Eingangsfrage nach den Selektionswirkungen des Review-Prozesses, so fragt sich, wie sich die in Tabelle 2 dargestellte Diskrepanz der Annahmequoten für Beiträge der qualitativen und quantitativen Sozialforschung erklären lässt. Eine abschließende Antwort ist nicht möglich, doch mögen einige Hinweise hilfreich sein. Zunächst ist zu konstatieren, dass für empirische Beiträge vorzugsweise Gutachtende ausgewählt werden, die für das jeweilige Forschungsgebiet einschlägig sind und/oder über die relevante methodische Expertise verfügen. Qualitative Studien werden also üblicherweise nicht abgelehnt, weil sie Gutachtenden vorgelegt würden, die der qualitativen Forschung fern stünden. Gelegentlich hört man aber die Ansicht, dass sich innerhalb des qualitativen Lagers verschiedene Schulen wechselseitig sehr kritisch begegneten. Eine solche Tendenz fällt in den Urteilen der Gutachten für die *ZfS* der Manuskripteingangsjahre 2015 und 2016 aber nicht auf: Die mittleren Urteile lagen für qualitative Beiträge (2,25) leicht über dem Gesamtschnitt von 2,15 (Skala 1 bis 4), für quantitative minimal darunter (2,12). Auch ist der Gutachterdissens im Spektrum der qualitativ orientierten Manuskripte geringer: Während sich die Gutachten dort in 17% der Fälle um zwei oder mehr Bewertungsstufen unterschieden, lag dieser Anteil im Bereich der quantitativen Beiträge bei 25%. Die Fachgutachtenden sind also kaum für die differenzielle Annahmequote verantwortlich zu machen.

Es waren eher die Herausgebenden, die Beiträge der qualitativen Forschung in ihren Voten kritischer beurteilten (1,51) als solche der quantitativen (1,75). Allerdings fielen die Urteile auch in dem Teil der Herausbergerschaft,

der der qualitativen Sozialforschung nahe steht, nicht positiver aus (1,53). Der Hauptgrund für die niedrige Annahemequote von Beiträgen der qualitativen Sozialforschung ist vielmehr in allgemeinen Defiziten wissenschaftlichen Arbeitens und Publizierens zu finden. Um dies zu untersuchen, habe ich die schriftlichen Begründungen der fünf Herausgebervoten für die 14 qualitativen Beiträge näher betrachtet, die ein Durchschnittsvotum von weniger als 1,30 erzielten – rund 40 % der Manuskripte dieser Gattung. Immer wieder wurden in den Votes folgende Mängel angeführt: Es fehle eine nachvollziehbare theoretische Rahmung des Aufsatzes; die Methoden des Sampling, der Datenerhebung bzw. -auswertung blieben intransparent; die Fragestellung sei unklar; es werde kein soziologisch relevantes Ergebnis erzielt; der Forschungsstand im jeweiligen Feld werde unzureichend dargestellt. Spezifische Aspekte qualitativer Forschung, etwa die Adäquanz der Methodenwahl oder die Konsistenz der Fallinterpretationen waren seltener der Stein des Anstoßes. Da es häufig Kumulationen dieser Probleme gab, erachteten die Herausgebenden eine Überarbeitung als aufwändig und wenig erfolgversprechend.

Derartige Probleme finden sich keineswegs nur in Beiträgen qualitativer Forschung, doch treten sie dort besonders häufig auf. Obwohl die qualitative Forschung in den letzten Jahrzehnten in der Soziologie einen klar erkennbaren Institutionalisierungsprozess vollzogen hat und personalstärker geworden ist, ist sie in den führenden allgemein-soziologischen Zeitschriften Deutschlands konstant wenig präsent (vgl. Kopp et al. 2012: 304). Aus meiner Sicht gibt es dafür systematische Gründe. Zum einen ist qualitative Forschung mit ihrer Materialfülle stark auf Buchpublikationen gerichtet, während die in der quantitativen Forschung vorzufindenden Trends zur Sekundärdatennutzung und zur kumulativen Promotion das Publizieren in Fachzeitschriften begünstigen. Entsprechend wird in den meist quantitativ ausgerichteten *Graduate Schools* das Publizieren in Journals trainiert. Dies erklärt den vergleichsweise geringen Manuskripteingang qualitativer Sozialforschung, nur bedingt jedoch deren Scheitern im Review-Prozess. Hinzu kommt aber zum anderen, dass sich viele qualitativ Forschende offenbar schwer damit tun, das konstitutive Merkmal der Offenheit des Forschungsprozesses in einen argumentativ stringenten Zeitschriftenartikel mit breiter Relevanz zu überführen. Wie in neueren Lehrbüchern der qualitativen Sozialforschung betont wird, benötigt eine (Zeitschriften-) Publikation eine klare Frage oder These, einen Abriss des Forschungsstandes, eine theoretische Rahmung, eine Offenlegung der Methodik und ein substanzielles Ergebnis (vgl. z. B. Przyborski & Wohlrab-Sahr 2010: Kap. 7). Auch

die Neubelebung der Gütekriterien-debatte geht von der berechtigten Einschätzung aus, dass es Qualitätsunterschiede in der qualitativen Sozialforschung gibt und dass Qualität intersubjektiv bewertbar sein muss (Strübing et al. 2018). Nach meiner Erfahrung sind es aber gar nicht zuvorderst Qualitätsstandards der qualitativen Forschung, die unterlaufen werden, sondern solche des wissenschaftlichen Arbeitens allgemein.

Die *ZfS* ist in ihrem Review-Verfahren bestrebt, suboptimale, aber viel versprechende Manuskripte gemeinsam mit ihren Autoren und Autorinnen produktiv weiter zu entwickeln und zur Publikation zu führen. Ein Mindestmaß der Grundanforderungen an einen Journalartikel muss aber von jeder Ersteinreichung erwartet werden, damit dies geschehen kann (vgl. dazu das Editorial in Heft 1 des Jahres 2017). Dies gilt für Manuskripte jeglicher Provenienz. Die *ZfS* lebt wie die Soziologie als Disziplin von der ganzen Breite an Methoden und Theorien, doch sollte der gemeinsame Nenner nachvollziehbaren, belastbaren wissenschaftlichen Publizierens stets Beachtung finden.

Literatur

- Bornmann, L., R. Mutz & H.-D. Daniel (2010): A Reliability-Generalization Study of Journal Peer Reviews: A Multilevel Meta-Analysis of Inter-Rater Reliability and its Determinants. In: *PLoS ONE* 5 (12): e14331.
- Hirschauer, S. (2004): Peer Review Verfahren auf dem Prüfstand. Zum Soziologiedefizit der Wissenschaftsevaluation. In: *Zeitschrift für Soziologie* 33(1): 62–83.
- Kopp, J., J. Schneider & F. Timmler (2012): Zur Entwicklung soziologischer Forschung. In: *Soziologie* 41: 293–310.
- Przyborski, A. & M. Wohlrab-Sahr (2010): *Qualitative Sozialforschung. Ein Arbeitsbuch*. 3. Auflage. München: Oldenbourg.
- Sahner, H. (1982): Zur Selektivität von Herausgebern: Eine Input-Output-Analyse der „Zeitschrift für Soziologie“. In: *Zeitschrift für Soziologie* 11(1): 82–98.
- Strübing, J., S. Hirschauer, R. Ayaß, U. Krähnke & T. Scheffer (2018): Gütekriterien qualitativer Sozialforschung. Ein Diskussionsanstoß. In: *Zeitschrift für Soziologie* 47(2): 83–100.
- Weller, C. (2004): Beobachtungen wissenschaftlicher Selbstkontrolle. Qualität, Schwächen und die Zukunft des *Peer Review*-Verfahrens. In: *Zeitschrift für Internationale Beziehungen* 11(2): 365–394.

Aus dem Herausbergremium

Nach vier Jahren sind turnusgemäß *Claudia Diehl* und *Gunnar Otte* aus dem Herausbergremium der *Zeitschrift für Soziologie* ausgeschieden. Sie hatten in dieser Zeit nicht nur das normale Pensum zu bewältigen, sondern

waren auch in viele Grundsatzentscheidungen involviert, die mit dem Verlagswechsel unserer Zeitschrift verbunden waren. Wir danken ihnen sehr für viele wichtige, mitunter kontroverse, aber immer anregende Diskussionen. Gunnar Otte sei an dieser Stelle zudem auch für das voranstehende Editorial zur Eröffnung des aktuellen Jahrgangs gedankt. In unser Herausbergremium neu gewählt wurden *Katrin Auspurg* (München) und *Karin Kurz* (Göttingen). Wir heißen sie herzlich willkommen und freuen uns sehr auf die Zusammenarbeit.

In der *Zeitschrift für Soziologie* wird es zudem eine weitere Neuerung geben, die wir auf Wunsch vieler Kolleginnen und Kollegen einrichten. Man könnte aber auch formulie-

ren: Wir greifen eine alte Tradition wieder auf, die insbesondere in den 1970er Jahren in unserer Zeitschrift verfolgt wurde. Neben den Standard-Beiträgen veröffentlichen wir zukünftig auch Forschungsskizzen, um Kolleginnen und Kollegen einen schnelleren Publikationszugang zu gewähren. Solche Forschungsskizzen beschränken sich auf nur wenige Seiten. Auch sie werden aber selbstverständlich einem sorgfältigen Begutachtungsprozess unterzogen. Weitere Angaben zu den formalen Voraussetzungen finden sich auf unserer Homepage.

Bielefeld, im Februar 2019

Die Herausgeberinnen und Herausgeber