

TMC Research

Evaluation des Peer Circle-Experiments 2022 der
Alexander von Humboldt-Stiftung

Peter van den Besselaar, Charlie Mom, Michelle Herte

TMC Research

Middenweg 203
1098 AN Amsterdam
Niederlande

E-Mail: info@teresamom.com
Telefon: +31 20 663 2184

(leere Seite)

Vorwort

Dieser Bericht ist das Ergebnis der Evaluierungsstudie zur Begleitung des von der Alexander von Humboldt-Stiftung 2022 durchgeführten Peer Circle-Experiments. Begutachtungsmodelle und ihre Auswirkungen sind ein wichtiges Thema, und die Ergebnisse dieses Berichts können dazu beitragen, besser zu verstehen, wie Begutachtungs- und Auswahlverfahren zur Fördermittelvergabe funktionieren und wie sie verbessert werden können.

Das Projekt begann im Frühjahr 2021 mit der Entwicklung des Forschungsplans, der Anfang 2022 angenommen wurde. Das Peer Circle-Experiment fand in zwei Runden statt: der Sommerrunde von März bis Juni 2022 und der Herbstrunde von Mitte August bis Mitte November 2022.

Die Autoren danken den Mitgliedern der Peer Circle (den Fachgutachter*innen) für ihre Mitwirkung an den Interviews und an zwei Umfragen. Wir bedanken uns außerdem bei den Ausschussmitgliedern und den AvH-Beschäftigten für die Teilnahme an Interviews und dafür, dass wir als Beobachter an den Ausschusssitzungen teilnehmen durften. Die Alexander von Humboldt-Stiftung hat die Studie finanziell unterstützt, Daten bereitgestellt und die Datenerhebung ermöglicht.

Dr. Michelle Herte (AvH) verfasste die Beschreibung des HFST-Förderinstruments und der Begutachtungsverfahren, die als Kapitel 2 und 3 in diesen Bericht aufgenommen wurden. Sie fungierte außerdem als Bindeglied zur AvH-Stiftung und arbeitete als solches schnell und präzise.

Amsterdam, den 12. Februar 2023

(leere Seite)

Managementfassung

In der letzten Zeit haben Diskussionen über die Qualität und die Nachhaltigkeit von Fachgutachten (für Förderungen) eine Reihe von Vorschlägen zur Verbesserung des Fachgutachtensystems ausgelöst. Das Hauptproblem besteht darin, dass die Gewinnung von Gutachter*innen sich als zunehmend schwierig erweist, da viele potenzielle Gutachter*innen keine Zeit haben. Allein die Alexander von Humboldt-Stiftung (AvH) benötigt zusätzlich zu den von Mitgliedern des Auswahlausschusses erstellten gutachterlichen Stellungnahmen jedes Jahr rund 5.000 externe Gutachten für etwa 3.200 Anträge. Für die AvH-Auswahlverfahren sind, einschließlich Gastgeber*innen-Stellungnahmen und Referenzen, jährlich rund 12.000 Fachgutachten und Stellungnahmen der Wissenschaftsgemeinde erforderlich. Die von der AvH überwachten Daten zeigen, dass etwa 60 % der Ersuchen um eine gutachterliche Stellungnahme, zumeist aus Zeitgründen, abgelehnt werden. Aufgrund dieser Belastung des Systems müssen alternative Ansätze in Erwägung gezogen werden. Gibt es andere, weniger zeitaufwändige Modelle zur Organisation des Begutachtungsverfahrens, die dieses Problem lösen würden? Ist dies ohne Beeinträchtigung der Qualität der Beurteilung möglich? Und würde die Wissenschaftsgemeinde ein solches alternatives Modell akzeptieren?

Im HFST-Programm – auf dem der Schwerpunkt dieses Berichts liegt – besteht das Ziel darin, für jeden Antrag zwei externe gutachterliche Stellungnahmen einzuholen. Beim klassischen Fachgutachtenmodell werden die Fachgutachter*innen für jeden Antrag individuell ausgewählt und angefragt. Das Peer Circle-Modell ist ein Versuch, das Fachgutachtensystem durch die Einführung eines kollaborativen Verfahrens zu verbessern. Bei diesem Modell besteht der Peer Circle (PC) aus fünf bis zehn Fachgutachter*innen sowie zusätzlichen Ausschussmitgliedern aus dem gleichen Fachgebiet, die zusammen zwischen 15 und 30 Anträgen in zwei Runden begutachten. Für jeden Antrag werden ein oder zwei der Fachgutachter*innen als „Erstgutachter*in“ benannt; sie beginnen das Begutachtungsverfahren mit einer strukturierten, aber umfassenden Beurteilung. Weitere Fachgutachter*innen begutachten die Teile der Anträge, zu deren Begutachtung sie sich angesichts ihrer Erfahrung und Fachkompetenz in der Lage sehen. Schließlich kommentieren die Fachgutachter*innen gegenseitig ihre gutachterlichen (Teil-)Stellungnahmen. Zusammengenommen sollte so eine hochwertige und vollständige Beurteilung des Antrags zustande kommen.

Das PC-Experiment wurde 2022 während der Sommer- und Herbstauswahlrunden des HFST-Programms in vier Fachgebieten durchgeführt: Anorganische Chemie, Materialwissenschaft, Zoologie und Neuere Geschichte. Zur Evaluierung des Experiments wurden diese Gebiete mit denselben Fachgebieten aus 2021, dem Jahr vor der Einführung des PC, und vier anderen Gebieten mit ähnlicher Fachkultur wie die vier Versuchsfachgebiete verglichen: Festkörperchemie, Werkstofftechnik, Pflanzenwissenschaft und Alte Geschichte.

Die Evaluierung soll die folgenden Fragen zum Peer Circle beantworten:

1. Ist die Qualität der gutachterlichen Stellungnahmen mindestens ebenso gut wie bei klassischen Stellungnahmen?
2. Führt die Interaktion im PC zu einem zu frühen Konsens, der eine kritische Betrachtung der Anträge verhindert?

3. Sehen die Fachgutachter*innen und Ausschussmitglieder den PC als bessere Alternative an?
4. Ist der PC effizienter als der klassische Ansatz?
5. Wählt der PC-Ausschuss die besten Antragsteller*innen aus?
6. Wirken sich PC-Stellungnahmen auf das Geschlechtergleichgewicht im Auswahlergebnis aus?

Daten und Methoden

Die Evaluierung beruht auf verschiedenen Datensätzen. Es wurden Interviews mit allen PC-Gutachter*innen und mit allen Ausschussmitgliedern, die an dem Experiment teilgenommen haben, sowie mit den beteiligten AvH-Beschäftigten durchgeführt. Die Fachgutachter*innen wurden außerdem gebeten, zwei Umfragen auszufüllen, eine nach der ersten Runde und eine nach der zweiten Runde. Diese Daten umfassten die Meinungen der PC-Teilnehmer*innen und eine Selbsteinschätzung des Begutachtungsverhaltens. Das Begutachtungsverfahren lief zum Großteil automatisiert auf einer Online-Plattform ab, so dass Protokolldateien von dieser Plattform zur Nachverfolgung des Begutachtungsverhaltens in Bezug auf Intensität und zeitliche Verteilung der Tätigkeiten genutzt werden konnten. Die Texte der gutachterlichen Stellungnahmen und der Anträge wurden analysiert und von der AvH bereitgestellte Verwaltungsdaten lieferten Informationen zu Antragsteller*in (Alter, akademisches Alter, Staatsangehörigkeit, institutionelle Zugehörigkeit usw.) sowie Informationen zur Bewertung. Schließlich wurden für zwei Fachgebiete bibliometrische Daten erhoben, um die akademische Leistung von geförderten und nicht geförderten Antragsteller*innen zu vergleichen.

Zusammenfassung der Ergebnisse:

1. Qualität und Umfang der PC-Stellungnahmen

Beinahe alle PC-Mitglieder sind der Ansicht, dass der PC umfassende gutachterliche Stellungnahmen erstellt, und wenn eine bestimmte Fachkompetenz fehlte, war es einfach, für bestimmte Themen einen externen Experten bzw. eine externe Expertin zum PC hinzuzuziehen. Der PC setzt sich aus unterschiedlichsten Fachgutachter*innen zusammen, die mehr Fachkompetenz einbringen, als dies bei klassischen gutachterlichen Stellungnahmen möglich ist. So entstehen noch bessere gutachterliche Stellungnahmen, da mehr Blickwinkel auf einen Antrag zu einer umfassenderen Stellungnahme, einer vollständigeren Abdeckung der verschiedenen Begutachungskriterien und zu einem transparenteren, objektiveren und weniger stark verzerrten Ergebnis führen. Die Beurteilung ist unabhängiger von der Meinung nur eines oder zweier Fachgutachter*innen. Ein weiterer Vorteil des PC besteht darin, dass die Fachgutachter*innen den Überblick über alle Anträge im PC haben und daher die Anträge eher vergleichend bewerten. Dies vereinfacht die Benotung und würde zu differenzierteren und realistischeren Noten führen.

Die meisten Ausschussmitglieder sehen auch Vorteile des PC gegenüber klassischen gutachterlichen Stellungnahmen. Dadurch, dass mehr Fachgutachter*innen kommentieren, steigt das Vertrauen der Ausschussmitglieder in die Ergebnisse der Beurteilung. Zudem sind die Argumente der Gutachter*innen (im Vergleich zu klassischen gutachterlichen

Stellungnahmen) klarer, und wenn etwas unklar ist, ist es leicht, sich zur Klärung an den*die Gutachter*in zu wenden. Der PC erleichtert auch den Ausschussmitgliedern die Arbeit, da sie in Fällen, in denen der PC zu einem Urteil kommt, das klar „fördern“ oder „ablehnen“ lautet, dieser Einschätzung folgen können, und sich so ausgehend von den in der PC-Beurteilung identifizierten kritischen Punkten auf die weniger klaren Fälle konzentrieren können. Dies kann funktionieren, da mehr Augen im PC zu einer höheren Sicherheit für die Ausschussmitglieder führen. Schließlich hat der Peer Circle den Vorteil, dass er das Problem der verspäteten gutachterlichen Stellungnahmen löst.

Insgesamt sind die meisten Teilnehmer*innen des Experiments der Ansicht, dass der PC zu besseren gutachterlichen Stellungnahmen führt als das klassische Begutachtungsverfahren, und nur wenige sind unsicher oder bevorzugen das klassische Verfahren.

2. Konsens

Den meisten befragten Fachgutachter*innen war die Möglichkeit einer Beeinflussung in einem Gruppenprozess wie dem PC bewusst, und die meisten Gutachter*innen gaben an, in gewissem Grad von anderen beeinflusst worden zu sein. Der Einfluss von anderen hilft jedoch, den eigenen Standpunkt zu reflektieren und zu entwickeln, und dies ist positiv, da das Ziel des Bewertungsverfahrens darin besteht, zu einem Konsens zu gelangen, was gefördert werden soll.

Einfluss ist nicht das Gleiche wie ein vorzeitiger Konsens, der die Gefahr der Unterdrückung etwaiger abweichender Ansichten birgt. Allen Befragten (Fachgutachter*innen, Ausschussmitglieder und Fächergruppenleitungen) war dieses Risiko stark bewusst, aber so gut wie niemand war der Ansicht, dass es eingetreten war. Wenn dies doch der Fall war, wurde es im PC ausdrücklich angesprochen.

3. Akzeptanz

Die Teilnehmer*innen des Peer Circle-Experiments waren einerseits nicht zufällig ausgewählt, und ihre Anzahl ist für eine Verallgemeinerung der Ergebnisse viel zu gering. Andererseits kamen sie aus unterschiedlichen Forschungsgebieten, so dass eine angemessene fachliche Vielfalt herrschte. Innerhalb dieser Grenzen lässt sich schlussfolgern, dass der PC im Allgemeinen als vielversprechender Ansatz für die Zukunft wahrgenommen und geschätzt wird. In der Umfrage und den Interviews äußerte sich eine Mehrheit der PC-Mitglieder positiv über den PC als Alternative zum klassischen Fachgutachtenmodell und gab an, er sei die Lösung für die Zukunft. Die Fachgutachter*innen gaben auch an, dass sie durch das Lesen der Beurteilungen anderer Gutachter*innen sowohl ihre Begutachtungs- als auch ihre Antragsfähigkeiten weiterentwickeln können. Nur drei (der zehn) Mitglieder des PC Neuere Geschichte zogen das klassische Begutachtungsmodell vor. Die überwiegende Mehrheit der Ausschussmitglieder teilte die positive Einschätzung der Fachgutachter*innen, wobei die Vertreter*innen des Bereichs Neuere Geschichte wieder kritischer waren.

4. Effizienz des PC-Verfahrens

Die meisten PC-Mitglieder gaben an, dass der PC Zeit spart und dass der Zeitaufwand pro Antrag aufgrund einer effizienteren Texterstellung geringer ist. Trotz des höheren Leseaufwands wird der gesamte Zeitaufwand für den PC als angemessen wahrgenommen. Für den einzelnen Fachgutachter bzw. die einzelne Fachgutachterin ist der PC gegebenenfalls mit (etwas) mehr Zeitaufwand verbunden als eine einzelne gutachterliche Stellungnahme auf die herkömmliche Weise, für die Gemeinschaft als Ganzes spart er viel Zeit. Beispielsweise wurden 89 Anträge von 27 PC-Mitgliedern (Fachgutachter*innen) bearbeitet, wozu im klassischen Verfahren 178 Fachgutachter*innen erforderlich gewesen wären (bei Einhaltung der Norm von zwei Gutachter*innen pro Antrag).

Die AvH konnte PC-Mitglieder aus einem größeren Pool potenzieller Fachgutachter*innen rekrutieren, wodurch dem PC auch relativ junge Fachgutachter*innen angehörten. Die Interviews und die Protokolldateien zeigen, dass die Zahl und die Qualität der Beiträge zur gutachterlichen Stellungnahme nicht altersabhängig sind.

Die vom PC-Ausschuss zu bearbeitenden Anträge lösten im Durchschnitt weniger Diskussionen aus als in anderen Ausschüssen, was darauf hindeutet, dass das PC-Begutachtungsverfahren Unsicherheit verringert und dazu führt, dass die Ausschusssitzungen mindestens gleich effizient sind.

5. Auswahlsergebnisse

Aus der Literatur ist bekannt, dass die Art und Weise, wie Begutachtungs- und Auswahlverfahren organisiert werden, Auswirkungen auf das Ergebnis hat. Entspricht die Qualität der im PC erfolgreichen Anträge der Qualität derer, die beim klassischen Verfahren erfolgreich waren? Es wurden zwei Aspekte analysiert. Zunächst wurden die Erfolgsquoten und die Durchschnittswertungen des Auswahl Ausschusses zwischen den PC-Gebieten und den Kontrollgebieten verglichen. Dann wurden nur im Bereich Chemie bibliometrische Kennzahlen für die PC- und die Kontrollgebiete verglichen. Die Ergebnisse der beiden Vergleiche ergaben keinen signifikanten Unterschied zwischen PC und klassischem Fachgutachten, was darauf hindeutet, dass der PC zu einem ebenso strengen Entscheidungsfindungsprozess führt.

Die bibliometrische Analyse zeigt, dass sowohl für die Anorganische Chemie (PC) als auch die Festkörperchemie (Kontrollgruppe) die Gesamtwerte der geförderten Antragsteller*innen nicht höher sind als die der abgelehnten, was darauf hindeutet, dass die Veröffentlichungsleistung bei der Entscheidungsfindung weniger wichtig ist als anhand von Beobachtungen aus Gutachtenberichten und Ausschusssitzungen angenommen. Aufgrund der geringen Anzahl an analysierten Anträgen wäre dies in einer größeren Untersuchung zu bestätigen.

6. Geschlechtergleichgewicht

Wirkt sich der PC auf das Geschlechtergleichgewicht in den Ergebnissen des Auswahlverfahrens aus? Die Erfolgsquote von Frauen unter den Antragsteller*innen, die im Peer Circle beurteilt wurden, ist mit der Erfolgsquote der in dieser Evaluierung berücksichtigten Frauen, die auf klassische Weise beurteilt wurden, vergleichbar.

Dieser Vergleich bezieht sich jedoch auf die Gesamtebene, aber auf Ebene der einzelnen Forschungsgebiete unterscheiden sich die Ergebnisse nach Gebiet und Jahr stark. Dabei kann es sich um (zufällige) Fluktuationen handeln, die mit den geringen Zahlen und der unterschiedlichen Qualität der Anträge zusammenhängen. Die Unterschiede bei den Erfolgsquoten können auch auf einen geschlechtsbezogenen Verzerrungseffekt hinweisen, aber dies kann noch nicht beantwortet werden und bedarf einer weiteren Analyse: Es werden größere Stichproben und Daten zur Leistung der Antragsteller*innen benötigt. Zum jetzigen Zeitpunkt lässt sich schlussfolgern, dass sich der Peer Circle neutral auf die Erfolgchancen von Frauen auswirkt, aber dies gilt es weiter zu untersuchen.

Inhaltsverzeichnis

Vorwort	3
Managementfassung	5
Teil I. Fall und Methoden	12
1. Einleitung	13
2. Das Auswahlverfahren für das Humboldt-Forschungsstipendienprogramm	14
2.1 Programm und Auswahlkriterien	14
2.2 Klassisches Auswahlverfahren	14
2.3 Das Peer Circle-Verfahren	16
3. Die Peer Circle-Pilotstudie	17
3.1 Der Peer Circle-Prozess	17
3.2 Zeitlicher Rahmen und technische Umsetzung der Pilotstudie	18
3.3 Zusammenfassung: Rollen bei klassischer Begutachtung und bei der Peer Circle-Begutachtung	19
4. Der Evaluierungsansatz	20
4.1 Ziel und Fokus	20
4.2 Ansatz	20
4.3 Methoden zur Datenerhebung	22
4.4 Daten	23
4.5 Terminologie	24
Teil II – Evaluierungsergebnisse	25
5. Qualität der gutachterlichen Stellungnahmen	26
5.1 Umfang	26
5.2 Tiefe und Detailgrad	27
5.3 Zahl der Fachgutachter*innen	31
5.4 Benotung	32
6. Entscheidungsfindung und Auswahlsergebnisse	33
6.1 Vorbereitung der Entscheidungsfindung	33
6.2 Erfolgsquoten	35
6.3 Qualität der Anträge	36
6.4 Vielfalt der ausgewählten Anträge	39
7. Aktivität der Fachgutachter*innen	41
7.1 Unterschiede zwischen den Fachgebieten?	41
7.2 Peer Circle-Aktivitäten im Zeitverlauf	41
7.3 Vorgehensweise und Integration der Peer Circle-Aktivitäten	45
7.4 Zeitaufwand, Aufwand, Motivation	48
7.5 Interaktion, Diskussion, Konsens	51
7.6 Interaktion zwischen Ausschussmitgliedern und Fachgutachter*innen	53
7.7 Vergleich mehrerer Anträge	54
7.8 Akzeptanz	54

8. Kontext und Implementierung des Verfahrens	56
8.1 Peer Circle-Mitglieder	56
8.2 Anonymität	56
8.3 Allgemeine Aspekte des Verfahrens	57
8.4 Funktionalität und Benutzerfreundlichkeit	61
Teil III – Schlussfolgerungen und Empfehlungen	63
9. Schlussfolgerung	64
10. Empfehlungen	69
Separate Anhänge	
A1 Daten und Methoden	
A2 Anregungen zur Plattform	

Teil I Fall und Methoden

1. Einleitung

Es ist notwendig, neue Begutachtungsmodelle für die Vergabe von Fördermitteln zu erproben, da die Funktionsweise des klassischen Fachgutachtensystems problematisch ist und es zunehmend schwieriger wird, ausreichend Fachgutachter*innen zu finden. In diesem Forschungsbericht evaluieren wir das Peer Circle-Projekt, ein Experiment mit einer neuen Form der Begutachtung, das von März bis November 2022 in der Alexander von Humboldt-Stiftung (AvH) durchgeführt wurde.

Die Evaluierung wurde als multimethodisches Forschungsprojekt ausgelegt und vergleicht den Peer Circle mit dem klassischen Peer Review-Verfahren zur Fördermittelvergabe. Multimethodisch bedeutet in diesem Zusammenhang, dass wir eine breitere Auswahl an Daten und Methoden als üblich nutzen und die üblichen Datenquellen, wie Dokumente, Interviews und eine Umfrage mit anderen Datenquellen, wie bibliometrischen Daten, Textdaten und Protokolldateien der Onlineplattform, auf der die Begutachtungsaktivitäten stattfinden, miteinander kombinieren.

Die Interviews lieferten umfangreiche Dateninputs, die manchmal in dieselbe Richtung weisen, sich jedoch hin und wieder (teilweise) widersprechen. Wir haben versucht, dies so gut wie möglich widerzuspiegeln, aber können natürlich nicht sämtliche Äußerungen aufnehmen – auch weil sich Äußerungen dann leicht einzelnen Befragten zuordnen ließen, was es zu vermeiden gilt.

Der Bericht ist folgendermaßen aufgebaut: Teil I des Berichts besteht (neben dieser Einleitung) aus drei Kapiteln: Kapitel 2 enthält eine Beschreibung des HFST-Förderprogramms, in dessen Rahmen das Peer Circle-Experiment stattfand. In Kapitel 3 werden das klassische Gutachtenformat und der Peer Circle ausführlicher beschrieben. Diese beiden Kapitel wurden von Michelle Herte (AvH) verfasst. In Kapitel 4 wird der für die Studie verwendete Ansatz, einschließlich der verwendeten Datentypen, erörtert.

Teil II besteht aus vier Kapiteln zu den Ergebnissen dieser Studie: Die Qualität der gutachterlichen Stellungnahmen (Kapitel 5), der Entscheidungsfindungsprozess und die sich daraus ergebende Auswahl (Kapitel 6), die Aktivitäten innerhalb der Peer Circle (Kapitel 7) sowie die Konzeption und Implementierung des Peer Circle-Verfahrens (Kapitel 8).

Teil III umfasst die Schlussfolgerungen (Kapitel 9) und Empfehlungen (Kapitel 10).

Ein separates Dokument wird die Anhänge beinhalten: einer mit weiteren Details zu den verwendeten Daten und Methoden und ein anderer mit den Empfehlungen zur Online-Plattform.

2. Das Auswahlverfahren für das Humboldt-Forschungsstipendienprogramm

2.1 *Programm und Auswahlkriterien*

Um die Peer Circle-Pilotstudie und die Ergebnisse ihrer Evaluierung besser einordnen zu können, fassen wir zunächst das reguläre Auswahlverfahren für das Humboldt-Forschungsstipendium (HFST) zusammen.

Mit Humboldt-Forschungsstipendien ermöglicht die Alexander von Humboldt-Stiftung hochqualifizierten Wissenschaftler*innen aus dem Ausland längere Forschungsaufenthalte in Deutschland. Bewerben können sich Wissenschaftler*innen aus allen Fachgebieten und Ländern. Das Programm dient zur Förderung des herausragenden wissenschaftlichen Talents exzellenter Forschender zu Beginn ihrer Karriere. Es richtet sich daher an Postdoktorand*innen (weniger als vier Jahre nach Abschluss der Promotion) und erfahrene Forschende (weniger als 12 Jahre nach Abschluss ihrer Promotion) mit überdurchschnittlichen Qualifikationen.

Die akademische Qualifikation von Bewerber*innen wird anhand der folgenden *Auswahlkriterien* bewertet:

- Akademischer Werdegang und bisherige wissenschaftliche Leistung (Mobilität, Zielstrebigkeit, fachliche Breite, wissenschaftliche Produktivität)
- Qualität der im Antrag genannten Schlüsselpublikationen (Originalität, Innovationsgrad, Eigenbeitrag im Falle von Mehrautorenpublikationen)
- Originalität und Innovationspotenzial des vorgeschlagenen Forschungsvorhabens (Bedeutung für die Weiterentwicklung des Fachs, überzeugende Wahl der wissenschaftlichen Methoden, Aussichten des Bewerbers bzw. der Bewerberin auf wissenschaftliche Entwicklung, klare Fokussierung und Umsetzbarkeit innerhalb des beantragten Förderzeitraums und an der gewählten Gasteinrichtung)
- Zukunftspotenzial des Antragstellers bzw. der Antragstellerin (wissenschaftliches Potenzial, weitere akademische Entwicklung, Karriereperspektiven)

2.2 *Klassisches Auswahlverfahren*

Das Auswahlverfahren ist vollständig digitalisiert. Die Anträge können jederzeit online eingereicht werden und die Entscheidung erfolgt vier bis sieben Monate später bei einer der drei jährlichen Sitzungen des Auswahlausschusses im März, Juli und November.

Die Humboldt-Stiftung prüft vor Beginn der Begutachtung alle Anträge auf Zulässigkeit und Vollständigkeit. Während der *Begutachtungsphase* wird jeder Antrag von zwei unabhängigen Fachwissenschaftler*innen beurteilt, die von der Humboldt-Stiftung anhand ihres Fachgebiets unter Ausschluss von Wissenschaftler*innen mit möglichen Interessenkonflikten gemäß den Richtlinien der Stiftung zum Umgang mit Befangenheiten ausgewählt werden. Die Fachgutachten werden online über ein standardisiertes Formular mit folgenden Elementen abgegeben:

- Vier Schlüsselfragen, die die Auswahlkriterien (s. oben) widerspiegeln und sowohl offen als auch durch Bewertung des Niveaus der Antragssteller auf einer Skala von „unterdurchschnittlich“ über „durchschnittlich“ und „klar überdurchschnittlich (beste 15 %)“ bis hin zu „Spitzengruppe (beste 5 %)“ zu beantworten sind.

- Eine Zusammenfassung der positiven und negativen Aspekte des Antrags.
- Die Gesamtempfehlung des Fachgutachters bzw. der Fachgutachterin an den Auswahlausschuss, von „Ablehnung“ über „Grenzfall“ bis hin zu „Forschungsstipendium“.
- Eine Frage, ob es Aspekte gibt, die gegen eine Förderung sprechen (z. B. mögliche Konflikte mit rechtsverbindlichen Grundsätzen der Wissenschaftsethik, Gefahr eines rüstungsrelevanten Technologietransfers im Sinne der gesetzlichen Bestimmungen usw.), die bei Bejahung zu erläutern sind.

Nach der Begutachtung fügt die Humboldt-Stiftung eine schriftliche Managementfassung jedes Antrags auf Grundlage seiner Haupteigenschaften und der gutachterlichen Stellungnahmen hinzu und schlägt eine Stipendienvergabe oder eine Ablehnung des Antrags vor. Die Managementfassung und die Referenzgutachten werden zu den Antragsunterlagen, die dem *interdisziplinären Auswahlausschuss* eingereicht werden, hinzugefügt; dieser Ausschuss setzt sich folgendermaßen zusammen:

- Fachgebundene Ausschussmitglieder: anerkannte Wissenschaftler*innen unterschiedlicher Fachrichtungen (= Fachvertreter*innen)
- Nicht-fachgebundene Ausschussmitglieder: stimmberechtigte Vertreter*innen öffentlicher und privater Geldgeber*innen; nicht stimmberechtigte Vertreter*innen anderer wissenschaftlicher (Förder-)Organisationen.

Die Teilnehmer*innen der Ausschusssitzungen werden in gleichgroße interdisziplinäre Gruppen aufgeteilt, um über die ihrer Gruppe zugewiesenen Anträge zu diskutieren und zu entscheiden. Vor jeder Ausschusssitzung weist die Humboldt-Stiftung jeden Antrag dem *Fachvertreter bzw. der Fachvertreterin* zu, der/die dem betreffenden Forschungsgebiet am nächsten steht und keinen Interessenkonflikt hat.¹ Die Fachvertreter*innen werden gebeten, eine schriftliche Stellungnahme zu jedem Antrag abzugeben und ihn in eine von drei Kategorien einzuordnen:

- (1) eindeutig negative Fälle, die ohne Diskussion abzulehnen sind („A“-Fälle, für „Ablehnung“)
- (2) diskussionswürdige Fälle („D“-Fälle, für „Diskussion“)
- (3) eindeutig positive Fälle, die ohne Diskussion zu bewilligen sind („S“-Fälle, für „Stipendium“)

Alle Ausschussmitglieder werden vor der Auswahl Sitzung über die Kategorien (A, S, D) informiert. Wenn ein Mitglied einer A- oder S-Einstufung widerspricht, wird der betreffende Antrag als D-Fall behandelt, der in der Ausschusssitzung zu besprechen ist. Alle A- und S-Fälle werden zu Beginn der Sitzung *en bloc* entschieden. Anschließend stellen die Fachvertreter*innen die ihnen zugewiesenen D-Fälle vor, die dann vom Ausschuss diskutiert werden. Schließlich stimmen die Ausschussmitglieder für jeden D-Fall mit 2 Punkten (für Förderung mit hoher Priorität), 1 Punkt (für Förderung) oder 0 Punkten (nicht für Förderung) ab. Die Humboldt-Stiftung führt die Abstimmungsergebnisse aller Gruppen in einer Rangliste zusammen. Anträge gelten als positiv entschieden, wenn ihnen mehr als die Hälfte der Abstimmenden zustimmt (1 oder 2 Punkte). Gibt es mehr positiv entschiedene Anträge als

¹ Ausschussmitglieder mit einem Fachgebiet, das dem Thema des Antrags sehr nahekommt, können außerdem gebeten werden, eines der zwei Fachgutachten zu erstellen.

verfügbare Stipendien, werden die Anträge entsprechend ihrer Rangfolge in den Abstimmungsergebnissen bewilligt.

2.3 *Das Peer Circle-Verfahren*

Der Peer Circle ist ein neuer Ansatz zur Organisation der Begutachtungsphase des Auswahlverfahrens, mit dem das Problem angegangen werden soll, dass es zunehmend schwieriger ist, Fachgutachter*innen zu gewinnen und zu halten. Die Idee hinter dem Verfahren ist die Umgestaltung der Begutachtung zu einer kollektiven Aufgabe. Die Fachgutachter*innen werden nicht jeweils für einen einzelnen Antrag berufen, sondern für einen längeren Zeitraum, in dem sie aufgefordert werden, alle Anträge in ihrem Forschungsgebiet zusammen mit anderen Expert*innen zu begutachten. Alle Fachgutachter*innen aus dem gleichen Gebiet bilden zusammen mit den entsprechenden Fachvertreter*innen einen Peer Circle. Die Peer Circle können je nach Forschungsgebiet aus 7 bis 10 Mitgliedern bestehen. Bei der Berufung von Referenzgutachter*innen auf Basis ihrer Fachkompetenz bemüht sich die Humboldt-Stiftung um Diversität innerhalb der Peer Circle. Im Idealfall decken die Mitglieder eines Peer Circle eine Vielzahl von Fachgebieten und akademischen Altersgruppen (ab Senior-Postdoktorand*in) ab und sind geschlechtlich ausgewogen.

Für die Pilotstudie 2022 wurden vier Forschungsbereiche über alle Fachgebiete hinweg ausgewählt, um den Peer Circle-Ansatz im Rahmen des HFST-Programms zu beurteilen: (1) Neuere Geschichte, (2) Zoologie, (3) Anorganische Chemie und (4) Materialwissenschaft. Das Auswahlverfahren insgesamt wurde möglichst ähnlich zum regulären Verfahren des Programms gehalten. Somit wurden die eingereichten Anträge in den ausgewählten Gebieten auf Zulässigkeit und Vollständigkeit geprüft und von unabhängigen Fachexpert*innen (in diesem Fall den Peer Circle-Mitgliedern) begutachtet. Die Entscheidung wurde dann von einem interdisziplinär zusammengesetzten Auswahlausschuss getroffen, wobei jedem Antrag ein Fachvertreter bzw. eine Fachvertreterin zugewiesen wurde.

Die an der Peer Circle-Pilotstudie beteiligten HFST-Ausschussmitglieder traten separat von der regulären Auswahl Sitzung zusammen. Zusätzlich zu den vier aktiven Fachvertreter*innen aus den vier beteiligten Forschungsgebieten wurde ein weiteres Mitglied für jedes Gebiet rekrutiert, um Fachdiskussionen innerhalb des ansonsten interdisziplinären Ausschusses zu ermöglichen. Die Peer Circle-Gruppe des HFST-Ausschusses bestand mit zwei Mitgliedern aus jedem Gebiet insgesamt aus acht Mitgliedern. Somit war sie in Bezug auf Größe und Interdisziplinarität mit den anderen, regulären HFST-Ausschussgruppen vergleichbar.

Die Fachvertreter*innen wurden aufgefordert, die ihnen zugewiesenen Anträge nur in eindeutig negative Fälle (A) und diskussionswürdige Fälle (D) einzuordnen. Sowohl das Abstimmungssystem im Ausschuss und die Auswahlkriterien als auch die Qualitätsstandards des HFST-Programms waren unverändert anzuwenden. Die Erfolgsquote entsprach basierend auf der Summe der in der Auswahlrunde verfügbaren Stipendien der der übrigen Anträge im HFST-Programm. Nach den Auswahl Sitzungen der Peer Circle-Gruppe und der übrigen HFST-Gruppen wurden die Abstimmungsergebnisse in einer Rangfolge für alle Anträge zusammengeführt.

3. Die Peer Circle-Pilotstudie

3.1 *Der Peer Circle-Prozess*

Der Peer Circle ist ein kollaboratives Begutachtungsverfahren, das auf einer interaktiven Online-Plattform durchgeführt wird. Auf der Plattform können die Gutachter*innen auf alle maßgeblichen Antragsunterlagen ihres Forschungsgebiets zugreifen. Für einzelne Anträge kann der Zugriff für Gutachter*innen mit möglichen Interessenkonflikten entsprechend den Richtlinien der Stiftung zum Umgang mit Befangenheiten beschränkt werden. Peer Circle-Gutachter*innen nehmen ihre Einschätzungen vor, indem sie direkt zu den Anträgen Kommentare abgeben. Die Kommentare sind für die anderen Mitglieder sichtbar, sie können darauf Bezug nehmen, Fragen stellen usw. Fachvertreter*innen können mit Beginn der Begutachtungsphase auch auf Anträge aus ihren Gebieten zugreifen und optional die Gruppe der Gutachter*innen um ihre Meinung zu bestimmten Aspekten und um Klarstellungen bitten und sich entsprechend ihrer eigenen Fachkompetenz selbst an der Begutachtung beteiligen.

Um bei den Diskussionen die Anonymität zu gewährleisten, wurde allen Gutachter*innen ein eindeutiges, nicht offengelegtes Pseudonym zugewiesen. Während der Begutachtungsphase können die Gutachter*innen nur von den Beschäftigten der Humboldt-Stiftung identifiziert werden. Nach Fertigstellung aller gutachterlichen Stellungnahmen wird der Zugriff der Gutachter*innen deaktiviert und die Pseudonyme werden aufgehoben, so dass den Ausschusmitgliedern die Identität der Gutachter*innen genauso bekannt ist, wie dies auch im regulären HFST-Verfahren der Fall ist.

Peer Circle-Mitglieder können sämtliche Aspekte des Antrags kommentieren, werden aber dazu aufgefordert, von der Humboldt-Stiftung als Kommentare bereitgestellte Schlüsselfragen zu beantworten. Diese Fragen basieren auf den HFST-Auswahlkriterien und entsprechen den Schlüsselfragen, die während des klassischen Begutachtungsprozesses gestellt werden. Die Stellungnahmen können sich jedoch vergleichend auf andere Anträge derselben Auswahlrunde beziehen. Die im klassischen Verfahren für jede Frage geforderte Benotung erfolgt nur in der von den Gutachter*innen abgegebenen Gesamtbeurteilung, bei der die Anträge von 1 (unterdurchschnittlich) über 2 (überdurchschnittlich) und 3 (beste 15 % weltweit) bis 4 (beste 5 % weltweit) eingestuft werden. Da die Peer Circle-Gutachter*innen mehrere Anträge gleichzeitig beurteilen, wird ihnen nahegelegt, ihre Gesamtnote auf einen Vergleich der Kohorte zu stützen.

Zum Anstoß der Beurteilung eines Antrags beruft die Humboldt-Stiftung ein Peer Circle-Mitglied, dessen Fachgebiet dem des Antrags am nächsten kommt, zum *Erstgutachter bzw. zur Erstgutachterin*. Durch die ausdrückliche Beauftragung einer Person mit der Einleitung des Begutachtungsprozesses soll der Einstieg erleichtert werden. Die Humboldt-Stiftung bemüht sich, die Zahl der Beauftragungen als Erstgutachter*in gleichmäßig auf die Mitglieder des Peer Circle zu verteilen. Alle Peer Circle-Mitglieder werden gebeten, alle Anträge aus ihrem Forschungsbereich zu begutachten und mit anderen Gutachter*innen zu diskutieren. Zur Berücksichtigung der persönlichen (Zeit-)Planung stand des ihnen grundsätzlich auch frei, vor dem/der ernannten Erstgutachter*in mit der Beurteilung des Antrags zu beginnen.

Bei seltenen Fällen von stark interdisziplinären Anträgen oder Anträgen mit ungewöhnlicher Spezialisierung, die vom Peer Circle nicht adäquat abgedeckt werden kann, kann die

Humboldt-Stiftung Mitglieder eines anderen Peer Circle oder externe Expert*innen zur Begutachtung hinzuziehen und so die vorhandene Lücke schließen.

Während der gesamten Begutachtungsphase übernehmen zwei Beschäftigte der Humboldt-Stiftung pro Peer Circle die bedarfsgemäße Überwachung und Moderation des Prozesses. Zu ihren Aufgaben gehören die Beantwortung von (formalen) Fragen, beispielsweise zu Anträgen oder Auswahlkriterien, die Aufforderung der Gutachter*innen zur Ergänzung oder Klarstellung von unvollständigen oder unklaren Stellungnahmen, die Überwachung auf Vollständigkeit (d. h., ob alle wesentlichen Aspekte vom Peer Circle angemessen beurteilt wurden) und bei Bedarf die Berufung zusätzlicher Gutachter*innen. Am Ende der Begutachtungsphase fügt die Fächergruppenleitung der Stiftung einen Kommentar zur Zusammenfassung der wesentlichen Aspekte der gutachterlichen Stellungnahmen hinzu und fordert dann den/die benannte*n Fachvertreter*in auf, seine/ihre abschließende Beurteilung und Empfehlung hinzuzufügen.

3.2 Zeitlicher Rahmen und technische Umsetzung der Pilotstudie

Die Pilotstudie zum Peer Circle-Verfahren wurde im Rahmen des HFST-Programms in zwei Auswahlrunden durchgeführt: Sommer und Herbst 2022, von März bis November 2022. Die Peer Circle-Ausschusssitzungen fanden jeweils eine Woche vor der regulären HFST-Ausschusssitzung statt: am 29. Juni bzw. am 3. November. Die Fristen (für Prüfungen der Förderfähigkeit, Fachgutachten, Stellungnahmen der Fachvertreter*innen usw.) beruhten mit leichten Anpassungen auf den regulären Programmfristen.

Die Gesamtdauer der an das reguläre Verfahren angepassten Peer Circle-Begutachtungsphasen, während derer die Gutachter*innen Zugriff auf die Anträge hatten und diese kommentieren konnten, betrug in der Pilotstudie 2022 bis zu 11 Wochen. Für jede Auswahlrunde wurden zwei interaktive zweiwöchige Begutachtungsphasen festgelegt, direkt nach Beginn bzw. unmittelbar vor dem Ende der gesamten Begutachtungsphase. Die interaktiven Begutachtungsphasen waren als Angleichung der Begutachtungsaktivität der Peer Circle-Mitglieder vorgesehen, um die Kommunikation auf der Online-Plattform durch eine mögliche Verkürzung der Antwortzeiten zu erleichtern.

Das Peer Circle-Verfahren wurde mithilfe einer in Deutschland gehosteten Online-Proofing-Software getestet. Die Implementierung der Peer Circle-Methode war von den technischen Möglichkeiten der Plattform abhängig, weshalb für einige Verfahrensdetails Kompromiss- oder Umgehungslösungen erforderlich waren, die zumeist die Peer Circle-Mitglieder nicht betrafen. Andere hatten Kompromisse bei den folgenden Aspekten zur Folge:

- Pseudonym: Bei den Pseudonymen der Gutachter*innen handelte es sich nicht um aussagekräftige Nicknames, sondern um vom System automatisch zugewiesene alphanumerische Codes
- Tagging: Das namentliche Tagging von Nutzer*innen ist eine Systemfunktion, die dazu dient, diese Nutzer*innen über direkt an sie gerichtete Kommentare oder Fragen zu benachrichtigen. Während der Pilotstudie war es nicht möglich, Nutzer*innen direkt mit ihrem Pseudonym zu taggen, so dass die Gutachter*innen einander nicht taggen konnten. Sie konnten jedoch auf das Pseudonym verweisen oder Beschäftigte

der Humboldt-Stiftung taggen, die wiederum das Peer Circle-Mitglied taggen und somit benachrichtigen konnten.

3.3 Zusammenfassung: Rollen bei klassischer Begutachtung und bei der Peer Circle-Begutachtung

Insgesamt sind an beiden Verfahren die gleichen Teilnehmer*innen beteiligt, aber ihre Rollen unterscheiden sich in einigen Aspekten, die unten zusammengefasst sind.

Rolle	im Peer Circle-Verfahren	im klassischen Verfahren
Antragsteller*in	<ul style="list-style-type: none"> • Reicht Antrag ein 	<ul style="list-style-type: none"> • Reicht Antrag ein
Unabhängige*r Fachgutachter*in	<ul style="list-style-type: none"> • Peer Circle-Mitglied • Wird für einen längeren Zeitraum berufen • Begutachtet mehrere Anträge pro Runde • Fungiert als Erstgutachter*in für Anträge, die seinem/ihrer Fachgebiet am nächsten kommen • Kann Kommentare aller Peer Circle-Gutachter*innen lesen und interagieren • Kann Peer Circle-Mitglieder nicht (namentlich) identifizieren 	<ul style="list-style-type: none"> • Wird für einen einzigen Antrag berufen • Kennt weder andere Gutachter*innen noch ihre Beurteilungen
Fachvertreter*in	<ul style="list-style-type: none"> • Kann sich an der Begutachtung und den Peer Circle-Diskussionen beteiligen (optional) • Gibt abschließende Beurteilung und Empfehlung (Ablehnung oder Diskussion) für zugewiesene Anträge ab • Präsentiert zugewiesene Anträge dem interdisziplinären Ausschuss • Kennt die Identität aller Gutachter*innen (nach Abschluss der Begutachtungsphase) 	<ul style="list-style-type: none"> • Keine Beteiligung an der Begutachtung (außer bei Nähe zum Fachgebiet) • Gibt abschließende Beurteilung und Empfehlung (Ablehnung, Diskussion oder Stipendium) für zugewiesene Anträge ab • Präsentiert zugewiesene Anträge dem interdisziplinären Ausschuss • Kennt die Identität aller Gutachter*innen
Beschäftigte*r der Humboldt- Stiftung	<ul style="list-style-type: none"> • Beruft Fachgutachter*innen für längeren Zeitraum (6-9 pro Peer Circle) • Bearbeitet Anträge • Moderiert Peer Circle in Begutachtungsphase • Fasst Begutachtungsergebnisse zusammen • Weist Anträge Fachvertreter*innen zu 	<ul style="list-style-type: none"> • Bearbeitet Anträge • Beruft Fachgutachter*innen einzeln (zwei pro Antrag) • Fasst Anträge und Begutachtungsergebnisse zusammen und beurteilt sie • Weist Anträge Fachvertreter*innen zu

4. Der Evaluierungsansatz

4.1 Ziel und Fokus

Der Peer Circle ist eine neue Methode zur Begutachtung von Anträgen auf Fördermittel durch eine Gruppe von Fachgutachter*innen, die in den Kapiteln 2 und 3 beschrieben wird. Ziel dieses Berichts ist die Evaluierung des Peer Circle (PC) im Vergleich zum klassischen Begutachtungsverfahren (CR). Mit dem PC soll das sich zunehmend verschärfende Problem gelöst werden, das darin besteht, Fachgutachter*innen zu gewinnen und zu halten. Ein Hauptgrund für die Ablehnung von Begutachtungseinladungen ist Zeitmangel, weshalb der PC gegenüber dem klassischen Begutachtungsverfahren Zeit sparen sollte. Darüber hinaus ist der PC so ausgelegt, dass auch weniger erfahrene Wissenschaftler*innen an dem Begutachtungsverfahren teilnehmen können, so dass die Zahl der potenziellen Fachgutachter*innen erheblich steigen würde. Schließlich sollte die Wissenschaftsgemeinschaft den Peer Circle als Verbesserung gegenüber dem klassischen Peer Review zur Fördermittelvergabe akzeptieren. Diese Ziele sollten außerdem ohne Abstriche bei der Qualität des Begutachtungsprozesses erreicht werden.

Die langfristige Bereitschaft zur Teilnahme am PC lässt sich anhand eines Experiments von kurzer Dauer nicht beurteilen, aber es war relativ einfach, PC-Mitglieder zu gewinnen, und die meisten von ihnen setzen ihre Tätigkeit für ein zweites Jahr fort. Es lässt sich jedoch herausfinden, ob die PC-Mitglieder dem Peer Circle-Begutachtungsverfahren positiv gegenüberstehen und ob sie es gegenüber der klassischen Begutachtungsmethode bevorzugen. Daraus ergeben sich mehrere Fragestellungen für die Evaluierung des Peer Circle-Prozesses und seiner Ergebnisse:

1. Sind die PC-Gutachten den klassischen Fachgutachten qualitativ mindestens ebenbürtig?
2. Führt die Interaktion im PC zu einer vorzeitigen Konsensbildung, die eine kritische Betrachtung der Anträge verhindert?
3. Sehen die Fachgutachter*innen und Ausschussmitglieder den PC als bessere Alternative an?
4. Ist der PC in Bezug auf die verschiedenen Aspekte, wie den Zeitaufwand, effizienter als der klassische Ansatz?

Schließlich gibt es Fragen zum Entscheidungsfindungsprozess nach dem PC, der durch die unterschiedlichen Begutachtungsverfahren beeinflusst werden kann:

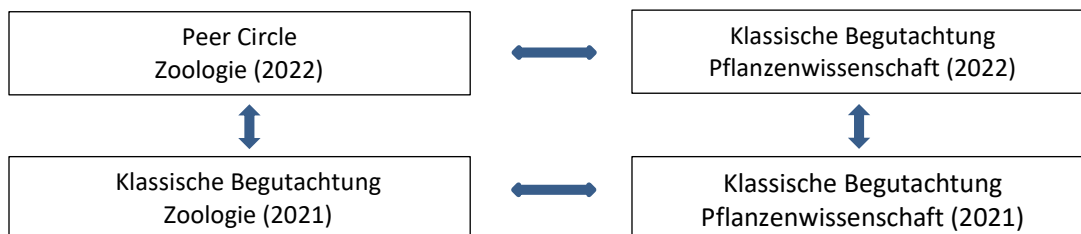
5. Wählt der PC-Ausschuss die besten Antragsteller*innen aus?
6. Wirken sich PC-Stellungnahmen auf das Geschlechtergleichgewicht im Auswahlresultat aus?

4.2 Ansatz

Zur Beantwortung dieser Fragen wurde der Peer Circle 2022 in zwei Runden für vier verschiedene Fachgebiete eingesetzt. Der PC wird mit dem klassischen Peer Review-Prozess in den gleichen Fachgebieten 2021 sowie mit ähnlichen Fachgebieten, die den klassischen

Begutachtungsprozess verwenden, 2022 und 2021 verglichen. Ähnlichkeit bezieht sich auf die Fachkultur, insbesondere deren Forschungs- und Bewertungsstil.

Für den Test des Peer Circle wurden die folgenden Fachgebiete ausgewählt: Anorganische Chemie, Materialwissenschaft, Zoologie (Biodiversität) und Neuere Geschichte. Vier Fachgebiete (Kontrollgruppen) wurden ausgewählt, weil sie den Versuchsfachgebieten am ähnlichsten sind: Festkörperchemie, Werkstofftechnik, Pflanzenwissenschaft und Alte Geschichte. Für den Peer Circle im Fachgebiet Zoologie sieht der Vergleich wie folgt aus:



Es wurden keine sozialwissenschaftlichen Fachgebiete einbezogen. Die Studie umfasst 325 Anträge und 28 Peer Circle-Mitglieder. Die nachfolgende Übersicht zeigt die Verteilung der Anträge auf die acht Fachgebiete und die zwei Jahre.

<u>Versuchsfachgebiete</u>	2021	2022
	Vor Peer Circle	Peer Circle
Anorganische Chemie	29	20
Materialwissenschaft	13	19
Zoologie (Biodiversität)	24	30
Neuere und Neueste Geschichte	32	20
<u>Kontrollfachgebiete</u>		
Festkörperchemie	21	14
Werkstofftechnik	16	21
Pflanzenwissenschaft	15	19
Alte Geschichte	17	15

Zwei Runden: Aus verschiedenen Gründen musste die Studie mindestens zwei Peer Circle-Runden umfassen. Erstens verhindern mehrere Runden mit denselben Peer Circle-Mitgliedern, dass wir nur „Neuheitseffekte“ beobachten, die zu einer positiven oder negativen Verzerrung führen können. Zweitens geben mehrere Runden den Peer Circle-Mitgliedern die Möglichkeit, sich an die Tools und die Arbeitsweise zu gewöhnen. So wird verhindert, dass die Evaluierungsergebnisse stärker die Lernkurve als die „stabile“ Nutzung der Peer Circle-Methode widerspiegeln. Drittens ist die Zahl der Anträge pro Fachgebiet in jeder Runde eher gering (10 bis 15), so dass mehrere Peer Circle-Runden einen größeren Datensatz erzeugen, der robustere Ergebnisse liefert. Es wurde entschieden, zwei Runden zu verwenden: die Sommerrunde und die Herbstrunde 2022.

Unterschiede zwischen den Fachgebieten: Vier unterschiedliche Fachgebiete wurden in die Studie einbezogen, da es möglich sein könnte, dass es beim Einsatz und der Akzeptanz neuer Formen der Begutachtung Unterschiede zwischen den Fachkulturen gibt. Wir werden die Interviews, Umfragen und sonstigen Daten auf deutliche Unterschiede zwischen den Fachgebieten prüfen, die auf solche Unterschiede hinweisen.

4.3 Methoden zur Datenerhebung

Es wurden unterschiedliche Methoden zur Datenerhebung eingesetzt. Bei einigen standen die Meinungen der Peer Circle-Mitglieder, der Ausschussmitglieder und der AvH-Beschäftigten im Vordergrund (Interviews, Umfrage), bei anderen ging es um Verhaltensdaten (Protokolldateien, Beobachtungen, schriftliche Kommentare/Stellungnahmen, bibliometrische Daten).

Beobachtungen: Der Peer Circle-Prozess wurde mithilfe von *Protokolldateien* der Online-Plattform beobachtet. Aus diesen Daten gehen die Zahl der Eingriffe durch die AvH-Beschäftigten, die Zahl der Beiträge von PC-Mitgliedern und Ausschussmitgliedern, der Grad der wechselseitigen Interaktion und die zeitliche Verteilung der Aktivitäten hervor. Die für den Begutachtungsprozess aufgewandte Zeit lässt sich anhand der Protokolldateien grob schätzen.

Die zwei PC-Ausschusssitzungen und die zwei regulären HFST-Ausschusssitzungen wurden beobachtet, um herauszufinden, ob es Unterschiede beim Ablauf der Sitzungen gab, die mit unterschiedlichen Begutachtungsmodellen, PC und klassische Begutachtung, zusammenhängen können.

Interviews: Die Peer Circle-Mitglieder, die Ausschussmitglieder und ausgewählte AvH-Beschäftigte wurden zu ihren Erfahrungen mit dem Peer Circle, zu den Vor- und Nachteilen sowie zu ihrer Zufriedenheit mit dem neuen Modell im Vergleich zur klassischen Form der Begutachtung im Hinblick auf den Prozess und die Qualität der Beurteilungen befragt. Die Befragten sollten auch die Erfahrungen mit dem Peer Circle mit ihren Erfahrungen mit der klassischen Begutachtung zur Fördermittelvergabe vergleichen. Es ist festzuhalten, dass die klassische Begutachtungsmethode im Rahmen des AvH-HFST-Programms, wie die Beschreibung in Kapitel 2 zeigt, stärker strukturiert ist, als dies allgemein der Fall ist: ein Frage-Antwort-Format. Die Befragten und die Umfrageteilnehmer*innen vergleichen jedoch den PC mit den gängigen Formen des Peer Review, bei denen die Fachgutachter*innen ihr eigenes Format für die Begutachtung wählen können.

Wir haben die Hälfte der 28 Fachgutachter*innen nach der ersten Runde und die andere Hälfte nach der zweiten Runde befragt. Wir haben acht der Ausschussmitglieder und vier der Fächergruppenleitungen zwei Mal (nach jeder Runde) und vier Sachbearbeitungen nach der ersten Runde befragt. Insgesamt wurden 56 Interviews durchgeführt. Es wurde eine Themenliste als Leitfaden für die Interviews erstellt. Die Ausschussmitglieder nahmen in der ersten Runde im Anschluss an die Ausschusssitzung auch an einer Gruppendiskussion teil. Die AvH-Beschäftigten wurden zu ihren Erfahrungen mit dem Peer Circle im Vergleich zum klassischen Ansatz befragt: Was lief gut, welche Probleme traten auf?

Umfrage: Die Ergebnisse der ersten Interviewrunde wurden zur Entwicklung der Umfrage genutzt, in der die gleichen Themen wie in den Interviews abgedeckt wurden. In der ersten Runde haben wir die Umfrage unter den PC-Mitgliedern, die in dieser Runde nicht interviewt wurden, durchgeführt, aber in der zweiten Runde des Peer Circle richtete sich die Umfrage mit einem leicht abgeänderten Fragebogen an alle Mitglieder; 25 (von 28) haben teilgenommen. Bei denjenigen, die zwei Mal an der Umfrage teilgenommen haben, konnten wir die Umfrageergebnisse miteinander vergleichen. Dies liefert einige Informationen zu Meinungen, die sich durch die zunehmende Vertrautheit mit dem PC möglicherweise ändern.

Gutachterliche Stellungnahmen und Kommentare: Die gutachterlichen Stellungnahmen der PC-Mitglieder bestehen aus Antworten auf eine Reihe von Fragen zum Antrag sowie aus Kommentaren anderer PC-Mitglieder zu diesen Stellungnahmen. Länge, Detailgrad und abgedeckte Bewertungsdimensionen werden mit der klassischen gutachterlichen Stellungnahme verglichen. Der Vergleich wird mithilfe der Interviews und der Umfrage (s. oben) sowie durch eine *Textanalyse* der gutachterlichen Stellungnahmen, die möglicherweise eine objektivere Analyse des Inhalts und Stils der Texte liefert, durchgeführt. Genauer werden die in den gutachterlichen Stellungnahmen verwendeten Wörter in linguistische Kategorien eingeordnet und die Häufigkeit der Nutzung dieser linguistischen Kategorien ermöglicht die Identifizierung der entsprechenden Eigenschaften der Texte der Stellungnahmen, wie Schreibstil und Fokus der gutachterlichen Stellungnahme.²

Die Entscheidungsdaten: Das Ergebnis des Auswahlprozesses ist eine Rangordnung der Antragsteller*innen sowie eine Gruppe ausgewählter Antragsteller*innen und eine Gruppe nicht ausgewählter Antragsteller*innen. Da die Produktivität und der Impact von Zeitschriftenartikeln in vielen Naturwissenschaften akzeptierte Auswahlkriterien sind, könnte man erwarten, dass die erfolgreichen Antragsteller*innen *im Durchschnitt* in Bezug auf Veröffentlichungen und Impact bessere Leistungen aufweisen. Die Bewertung der Antragsteller*innen beruht nicht nur auf ihrer bisherigen Veröffentlichungsleistung, sondern auch auf der Qualität des Vorhabens, dem Lebenslauf, der Mobilität und dem Zukunftspotential der Antragsteller*innen. Man würde dennoch eine positive Korrelation zwischen der bisherigen Leistung und der wahrgenommenen Qualität des Antragstellers bzw. der Antragstellerin erwarten.³ Wir haben *bibliometrische Daten* erhoben, die Indikatoren⁴ berechnet, und vergleichen das bisherige durchschnittliche Leistungsniveau der abgelehnten und der angenommenen Anträge. Dies wurde im Bereich Chemie für vier Gruppen durchgeführt: den Peer Circle Anorganische Chemie und die drei Kontrollgruppen Anorganische Chemie 2021 sowie Festkörperchemie 2021 und 2022. Das Ziel ist eine erste Antwort auf die Frage nach der Qualität der Antragsteller*innen und der ausgewählten Anträge und die Beurteilung des Geschlechtergleichgewichts im Ergebnis.

4.4 Daten

Zusammenfassung der verwendeten Datensätze:

1. Thematisch geordnete Interviewprotokolle. Diese umfassen Meinungen zu einer Reihe von Themen und Einschätzungen des eigenen Verhaltens. Sie wurden um Daten aus dem Protokoll einer Gruppendiskussion der Ausschussmitglieder nach der ersten Runde ergänzt.
2. Die Ergebnisse der Umfrage unter den PC-Mitgliedern zur Einschätzung des eigenen Verhaltens und den Meinungen zu einer Reihe von Themen.

² Weitere Einzelheiten finden sich in Abschnitt 5.2.

³ Beispiel einer umfassenden Analyse: Van den Besselaar P & Mom C (preprint 2020), *Gender bias and grant allocation – a mixed picture*.

⁴ Wir verwenden die Zahl der Veröffentlichungen, die fraktionierte Berechnung der Veröffentlichungen, den *Source Normalized Impact per Paper* und die Zahl der am meistzitierten Veröffentlichungen (beste 1 % und beste 5 %). Die fraktionierte Berechnung bedeutet, dass die Zahl der Mitautor*innen berücksichtigt wird: Eine Veröffentlichung mit vier Autor*innen wird als $\frac{1}{4}$ gezählt. Weitere Einzelheiten finden sich in Abschnitt 6.2.

3. Protokolldateien: Zahl der Logins, Online-Zeit, aktive Zeit, Zahl und Datum der Gutachtenstellungen und Kommentare.
4. Beobachtungsdaten zu den vier beobachteten Ausschusssitzungen: Für jeden Antrag aufgewandte Zeit, Zahl der Diskussionsteilnehmer*innen für jeden Antrag.
5. Gutachtendaten: Textdaten aus den gutachterlichen Stellungnahmen und den Kommentaren für die einzelnen Fachgutachter*innen und je Begutachtungsdimension (vier Schlüsselfragen, Zusammenfassung) sowie Daten zu den von den Fachgutachter*innen vergebenen Wertungen.

Die Datensätze wurden um Daten zu den Antragsteller*innen ergänzt:

6. Die den Antragsteller*innen vom Ausschuss zugeordneten Wertungen und die Entscheidung (aus dem Dokument mit der Rangordnung, die sich aus der Abstimmung in den Ausschüssen ergibt).
7. Persönliche Eigenschaften (aus den Antragsunterlagen): Alter, Staatsangehörigkeit, Geschlecht, Wohnort, institutionelle Zugehörigkeit, Gasteinrichtung, aktuelle Position.
8. Bibliometrische Daten zur Messung von Produktivität und Impact der Antragsteller*innen im Bereich Chemie.

4.5 Terminologie

Es sind folgende Arten von Teilnehmer*innen in den Peer Circles zu unterscheiden:

- (i) Die Fachgutachter*innen, auch PC-Mitglieder genannt.
- (ii) Die Fachvertreter*innen, hier Ausschussmitglieder genannt.
- (iii) Beschäftigte der Alexander von Humboldt-Stiftung, die Fächergruppenleitung und die Sachbearbeitung. Beide Gruppen zusammen werden als AvH-Beschäftigte bezeichnet.

Verwendete Personalpronomen schließen alle Geschlechter ein.

Teil II – Evaluierungsergebnisse

5. Qualität der gutachterlichen Stellungnahmen

Was als qualitativ gutes Fachgutachten gilt, hängt von den Zielen des Förderprogramms ab. Da es unterschiedliche Arten von Förderungen gibt, unterscheiden sich auch die anzuwendenden Kriterien. Beim klassischen *Peer Review*-Verfahren werden jedoch in der Regel Fachgutachter*innen mit Kernkompetenzen im spezifischen Bereich des Förderantrags gesucht. Die PC-Mitglieder haben die Frage der geeigneten Kriterien in den Interviews angesprochen. Wenn ein Themenexperte bzw. eine Themenexpertin einen Antrag begutachtet, konzentriert sich die Stellungnahme häufig stärker auf den Inhalt des Forschungsplans als auf andere relevante Dimensionen. Die Frage ist dann, ob detaillierte fachliche Kommentare – die als Feedback hilfreich sein können – auch für den Auswahlprozess sehr relevant sind. Dieses Thema ist hier von Bedeutung, da im HFST-Programm der Schwerpunkt mehr auf den Leistungen und dem Zukunftspotenzial der einzelnen Antragsteller*innen als auf dem spezifischen vorgeschlagenen Projekt liegt, das gut sein sollte. Es wurde häufig angeführt, dass der PC möglicherweise eine bessere Methode zur Beurteilung der HFST-Anträge ist, da sich der breitere PC natürlich auf andere Aspekte als den fachlichen Inhalt des vorgeschlagenen Projekts konzentriert.

Inwieweit unterscheiden sich PC-Gutachten von den klassischen Fachgutachten? Es sollte darauf hingewiesen werden, dass die Qualität klassischer gutachterlicher Stellungnahmen erheblich variiert, so dass „die“ Qualität des klassischen Fachgutachtens kein Standard zur Bewertung der PC-Gutachten sein kann. Daher konzentrieren wir uns auf die Meinungen der PC-Teilnehmer*innen, die in den Interviews und Umfragen unter anderem zur Qualität der PC-Gutachten befragt wurden. Da die meisten PC-Mitglieder in unterschiedlichen Settings begutachtet haben, sind sie eine gute Quelle für Vergleiche zwischen dem PC und klassischen *Peer Review*-Verfahren. Wir ergänzen hierzu eine Textanalyse der gutachterlichen Stellungnahmen, um in einem ersten Schritt zu untersuchen, wie diese zur Evaluierung der Begutachtungsverfahren und -qualität beitragen kann.

5.1 *Umfang*

Beim klassischen *Peer Review* sollen die Fachgutachter*innen alle Aspekte des Antrags behandeln. Im PC werden nur ein oder zwei Mitglieder, deren Kompetenz dem Thema des Antrags nahekommt, als *Erstgutachter*innen* für den betreffenden Antrag ernannt und von ihnen wird erwartet, dass sie alle Evaluierungsdimensionen behandeln. Die übrigen PC-Mitglieder fügen ihre Kommentare zu Teilen des Antrags hinzu und können auf Kommentare der anderen Mitglieder reagieren. So sollte die PC-Beurteilung umfassend sein. Tatsächlich haben die PC-Mitglieder das Gefühl, dass die Antworten auf die Begutachtungsfragen und die Kommentare zusammen eine umfassende gutachterliche Stellungnahme ergeben, und nur zwei der 23 PC-Mitglieder, die diese Frage beantwortet haben, äußerten sich leicht negativ (Abb. 1).

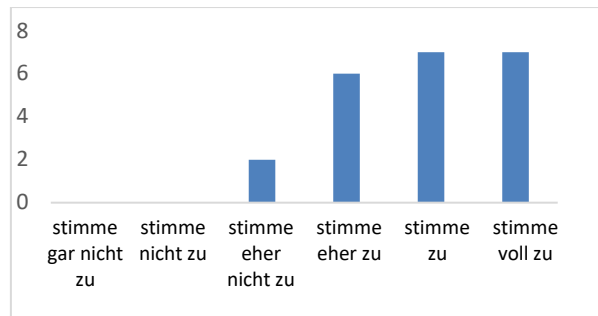


Abbildung 1: Ergeben die Kommentare zusammengenommen eine umfassende Einschätzung der Anträge?
N= 23

Auch in den Interviews der AvH-Beschäftigten findet sich diese positive Sichtweise. Beim klassischen Peer Review-Modell werden zwei Fachgutachter*innen ausgewählt, deren Fachgebiet dem vorgeschlagenen Projekt möglichst nahekommt. Im PC ist dies nicht der Fall, da die PC-Mitglieder ausgewählt werden, ehe die Anträge eingehen. Dies kann sich auf die Vollständigkeit des PC-Gutachtens auswirken, da im PC möglicherweise die Fachkompetenz fehlt. Die Interviews zeigen, dass die PC-Mitglieder dies als mögliches Problem sehen, aber gleichzeitig geben sie an, dass dieses Problem nicht aufgetreten sei.

- Erstens wurde bei wenigen Fällen aufgrund einer spezifischen Fachkompetenz ein externer Fachgutachter bzw. eine externe Fachgutachterin in den PC einbezogen und dies hat gut funktioniert.
- Zweitens sind beim HFST-Programm die Qualität, der Lebenslauf und die Zukunftsaussichten der Antragsteller*innen die Hauptkriterien und nicht nur oder hauptsächlich das Forschungsprojekt. Wie Fachgutachter*innen betonten, besteht die Gefahr, dass sich die Fachgutachter*innen, wenn ihr Fachgebiet dem Forschungsgegenstand sehr nahekommt, zu sehr auf die fachlichen Details konzentrieren, die häufig nicht sehr relevant für den Entscheidungsfindungsprozess sind.
- Drittens bilden die PC-Mitglieder zusammen einen großen Pool an kollektivem Wissen. Dies betrifft nicht nur die fachlichen Aspekte des Forschungsplans, sondern auch die Herkunftsländer der Antragsteller*innen. In einem PC ist die Wahrscheinlichkeit viel höher, dass eines der Mitglieder über Erfahrung im Herkunftsland verfügt und daher den Lebenslauf und die Karriere der Antragsteller*innen im lokalen Kontext besser beurteilen kann.
- Schließlich hoben einige der befragten Fachgutachter*innen und Ausschussmitglieder hervor, dass der Abstand zum fachlichen Inhalt eines Antrags kein PC-spezifisches Thema sei, da auch bei der klassischen Methode die „wissenschaftliche Nähe“ der Fachgutachter*innen nicht garantiert sei.

5.2 *Tiefe und Detailgrad*

In den Interviews äußerten viele Fachgutachter*innen die Ansicht, dass Tiefe und Detailgrad dem klassischen Fachgutachten mindestens ebenbürtig seien. Dies trifft vermutlich nicht auf alle Kommentare einzeln betrachtet zu, aber die größere Zahl an Kommentaren gleicht das aus. Und je mehr Personen sich mit einem Antrag befassen, desto objektiver wird die Beurteilung und desto besser wird das Gesamturteil.

Detailgrad und Tiefe können mit der Länge der gutachterlichen Stellungnahmen zusammenhängen, bei der es Unterschiede zwischen dem klassischen Peer Review-Verfahren und dem PC-Review gibt. In Tabelle 1 wird die Länge der gutachterlichen Stellungnahmen im Peer Circle mit der im klassischen Peer Review-Verfahren verglichen.⁵ Die Texte der gutachterlichen Stellungnahmen zu einem Antrag wurden zusammengefasst, und wir beschränken uns hier auf die Texte zum Lebenslauf, zu den Schlüsselpublikationen, zum vorgeschlagenen Forschungsprojekt und zum Zukunftspotenzial. Tabelle 1 zeigt die Durchschnittswerte und den Variationskoeffizienten (CoV) als Maß für die Streuung.

Tabelle 1: Textlänge der gutachterlichen Stellungnahmen⁶

		Summe*	Lebenslauf	Schlüssel- publikationen	Projekt- vorschlag	Zukunfts- aussichten
Klassisch	Durchschnitt	1253	347	292	422	192
	CoV**		0,48	0,59	0,52	0,52
Peer Circle	Durchschnitt	751	200	141	311	99
	CoV		0,67	0,85	0,59	0,70
Verhältnis	Klassisch/PC	1,67	1,74	2,07	1,36	1,94

* Summe der Wortzahl der vier Hauptteile der gutachterlichen Stellungnahme. Die Zusammenfassungstexte und die Profiltexthe haben wir nicht berücksichtigt, da sie im PC-Verfahren nicht existieren.

** Variationskoeffizient

Die klassischen gutachterlichen Stellungnahmen sind zusammen länger als das PC-Ergebnis und nach Angaben von Befragten kann dies möglicherweise daran liegen, dass mehrere Standardteile der klassischen gutachterlichen Stellungnahme, wie eine Zusammenfassung des vorgeschlagenen Projekts und Höflichkeitsfloskeln, fehlen. Einige der Befragten führten an, dass auch die klassischen gutachterlichen Stellungnahmen kürzer würden.

Die Erwartung war, dass dem Projektvorschlag im PC weniger Aufmerksamkeit zuteilwürde als im klassischen Peer Review-Verfahren. Somit wäre also zu erwarten, dass sich im PC, im Vergleich zum klassischen Peer Review, ein kleinerer Teil des Textes mit dem geplanten Projekt befasst. Dies ist jedoch nicht der Fall. In den klassischen Fachgutachten befassen sich 34 % des Textes mit dem Projektvorschlag, im Peer Circle sind es 41 %. Eine mögliche Erklärung hierfür ist, dass im PC viele Wörter verwendet werden, um das Projekt PC-Mitgliedern zu erklären, die mit dem Thema weniger gut vertraut sind.

Der PC wendet außerdem relativ gesehen weniger Wörter für die Schlüsselpublikationen und in etwa den gleichen Textanteil für die Besprechung des Lebenslaufs und der Zukunftsaussichten auf.

Einige Befragte gaben an, dass der *Schreibstil*⁷ in den PC-Gutachten und -Kommentaren informeller sei als bei der klassischen Peer Review-Methode, einschließlich der strukturierten

⁵ Wir unterscheiden hier nicht zwischen den beiden Runden, da die Länge in beiden Runden etwa gleich war.

⁶ In unseren Vergleichsgebieten, in denen das klassische Peer Review-Verfahren eingesetzt wurde, gab es 235 Antragsteller*innen, und bei rund 100 gab es nur einen externen Fachgutachter bzw. eine externe Fachgutachterin. Dies zeigt, wie schwierig es ist, genügend Fachgutachter*innen zu finden. In manchen dieser Fälle wurde das Ausschussmitglied, wenn es die passende Spezialisierung hatte, zum Zweitgutachter bzw. zur Zweitgutachterin berufen. In den übrigen Fällen haben wir für die Analyse die Aussagen aus der Stellungnahme des Ausschussmitglieds verwendet, auch wenn das Mitglied formal nicht Fachgutachter*in war. Die Zusammenfassungen der AvH-Beschäftigten wurden nicht für die Textanalyse verwendet.

Form der klassischen HFST-Begutachtung. Eine erste Analyse der Texte der gutachterlichen Stellungnahme bestätigt dies:

- Gutachten weisen normalerweise einen stark *analytischen Stil* (gemessen auf einer Skala von 0 bis 100) im Gegensatz zu einem eher narrativen Stil auf, was auch für die Fachgutachten für HFST gilt. Erwartungsgemäß ist der Wert für den analytischen Schreibstil bei PC-Gutachten etwas geringer: 87 gegenüber 97, und der Unterschied ist signifikant: $F(1, 322) = 89,87$, $p < ,000$.
- Ein weiterer Unterschied ist der Wert bei *Schlagkraft*; hier gilt, je höher der Wert, desto stärker betont der Fachgutachter bzw. die Fachgutachterin Fachkompetenz und Selbstvertrauen. Der PC weist mit 46 gegenüber 61 einen wesentlich geringeren Wert auf, was ebenfalls signifikant ist. $F(1, 322) = 159,30$, $p < ,000$.
- Das Gegenteil gilt für den Wert *authentisch*, der angibt, ob Texte offen bzw. aufrichtig im Gegensatz zu vorsichtig bzw. zurückhaltend formuliert sind: 48 für den PC gegenüber 26 für die klassischen gutachterlichen Stellungnahmen: $F(1, 322) = 99,14$, $p < ,000$.

Offenkundig wird für die Bewertung im PC ein anderer Schreibstil verwendet. Wie es eine der befragten Personen ausdrückte: „die Kommentare sind offen und ehrlich“. Eine andere bezeichnete es als „eine weniger stark kuratierte Begutachtung, weniger Bemühungen um eine kohärente Argumentation, die zum gewählten Urteil führt“.

Eine weitere Frage, für die mithilfe der Analyse der Gutachtentexte ein Antwortversuch vorgeschlagen werden kann, ist, ob die Fachgutachten der klassischen Methode ähnliche Schwerpunkte auf die verschiedenen Beurteilungskriterien legen wie die im Peer Circle. Mithilfe eines Wörterbuchs aus Begriffen, die sich auf die Beurteilungskriterien beziehen (in Bezug auf Werdegang, Mobilität, Eigenständigkeit, Veröffentlichungen, wissenschaftliche Leistung, Projekt, Qualität der Gasteinrichtung und Exzellenz), haben wir die Gutachtentexte des PC mit denen der klassischen Peer Review-Methode verglichen. Wie in Tabelle 2 zu sehen, gibt es einige Unterschiede. Erstens weisen die PC-Gutachten einen signifikant höheren Anteil an *gemeinsprachlichen Wörtern* auf, was darauf hinweist, dass sich in diesen gutachterlichen Stellungnahmen weniger Fachbegriffe finden. Dies passt zu der Erwartung, dass im PC die Aufmerksamkeit weniger auf den fachspezifischen Details und mehr auf den übrigen Beurteilungskriterien liegen würde, was sich aus den Interviews ergab. Dies entspricht auch der Beobachtung, dass die klassischen gutachterlichen Stellungnahmen signifikant mehr projektbezogene Wörter enthalten. Beim klassischen Gutachten schenken die Stellungnahmen der Leistung mehr Aufmerksamkeit als beim PC-Gutachten, bei den PC-Gutachten stehen hingegen die Gasteinrichtungen und die Unabhängigkeit der Antragsteller*innen stärker im Fokus.

Eines der Kriterien ist die Qualität der Schlüsselpublikationen. Aus Interviews mit mehreren PC-Mitgliedern geht hervor, dass sie die Schlüsselpublikationen überfliegen und nicht lesen, und die Fächergruppenleitungen haben dieses Thema auch angeführt. Die Diskussion über Veröffentlichungen dreht sich mehr um Anzahl, Zeitschriften-Impact und Zitierungen als um Inhalt und Ergebnisse. Stellt dies einen Unterschied gegenüber dem klassischen Peer Review-Verfahren dar, bei dem für jeden Antrag Spezialist*innen ausgewählt werden? Unsere Daten

⁷ Van den Besselaar P, Sandström U, Schiffbaenker H (2018), Using linguistic analysis of peer review reports to study panel processes. *Scientometrics* 117, 313-329; Van den Besselaar P, & Mom C (2022), The effect of writing style on success in grant applications. *Journal of Informetrics* 16(2):101257

weisen darauf hin, dass dies nicht der Fall ist, und die Häufigkeit der Verwendung von „bibliometrischen“ Begriffen in klassischen gutachterlichen Stellungnahmen ist mit der in den PC-Gutachten vergleichbar (Tabelle 2). Die Beobachtung der Diskussionen in den Auswahlausschüssen führt zur gleichen Schlussfolgerung, und auch dort werden die quantitativen Merkmale der Publikationslisten häufig erwähnt.

Tabelle 2: Unterschiede beim Beurteilungsschwerpunkt zwischen PC-Review und klassischem Peer Review

	Peer Circle-Review	Klassischer Peer Review	Einfaktorielle ANOVA
Gemeinsprachliche (nicht fachsprachliche) Wörter	80,0 %	76,0 %	$F(1, 322) = 46,03, p = ,0000$
Werdegang (inkl. Mobilität)	0,93 %	0,96 %	n.s.
Leistung (inkl. Schule & Hochschule)	0,60 %	0,72 %	$F(1, 322) = 5,03, p = ,0257$
Veröffentlichungen (Bibliometrie)	1,26 %	1,17 %	n.s.
Vorgeschlagenes Projekt	1,04 %	1,24 %	$F(1, 322) = 9,14, p = ,0027$
Eigenständigkeit	0,07 %	0,02 %	$F(1,322) = 31,52, p = ,0000$
Exzellenz	1,39 %	1,47 %	n.s.
Gasteinrichtung	0,23 %	0,17 %	$F(1, 322) = 7,00, p = ,0086$
Abschließende Wertung des Ausschusses	0,55 %	0,58 %	n.s.

n.s = nicht signifikant

Im Allgemeinen befinden die Befragten Tiefe und Umfang der PC-Gutachten für gut; nur der fachliche Inhalt der Anträge werde bei den klassischen Fachgutachten stärker abgedeckt. Mehrere Befragte begrüßen den geringeren Fokus auf fachspezifische Details, da dies verhindert, dass fachliche Details bei der Begutachtung und im Auswahlverfahren zu viel Bedeutung einnehmen. Dies wird durch die Umfrage gestützt, da nach Ansicht beinahe aller PC-Mitglieder das Gesamtergebnis der Begutachtung eine hohe Qualität aufweist (Abb. 2) und darüber hinaus halten es siebzehn von zwanzig für mindestens genauso gut wie das klassische Peer Review-Verfahren (Abb. 3).

Zehn PC-Mitglieder halten ihre eigenen Beiträge für (geringfügig) weniger gut als im klassischen Verfahren, während nur sieben sie als (etwas) besser ansehen (Abb. 4). Es ist darauf hinzuweisen, dass sechs PC-Mitglieder diese Frage nicht beantwortet haben, wahrscheinlich weil sie kaum Begutachtungserfahrung hatten und daher keinen Vergleich anstellen konnten. Jedoch sind, wie eine der befragten Personen äußerte, die Kommentare des PC zusammen dennoch besser („Schwarmintelligenz“), auch wenn die einzelnen Beurteilungsbeiträge nicht so gut sind wie im klassischen Verfahren.

Zusammengefasst halten die PC-Mitglieder die gutachterlichen Stellungnahmen/Kommentare für dem klassischen Verfahren mindestens ebenbürtig. Diese Meinung teilen auch die Fächergruppenleitungen, die auch in Bezug auf Umfang, Tiefe und Detailgrad eine positive Einschätzung vertraten. Gleiches gilt für die meisten Ausschussmitglieder, die die Argumente der Fachgutachter*innen teilen.

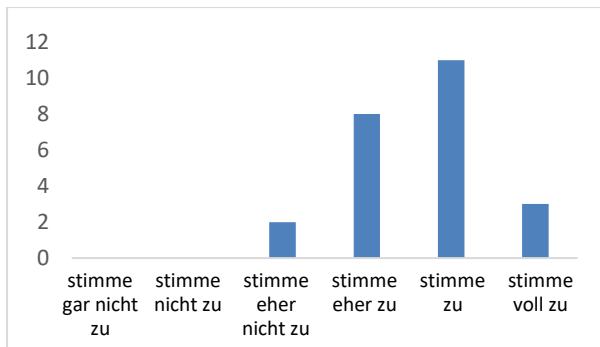


Abbildung 2: Die Kommentare und Beurteilungen sind von hoher Qualität
N = 24

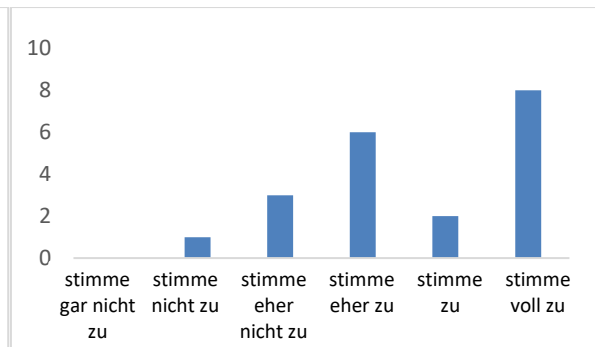


Abbildung 3: Die Qualität der PC-Beurteilung entspricht mindestens der von klassischen Gutachten
N = 21

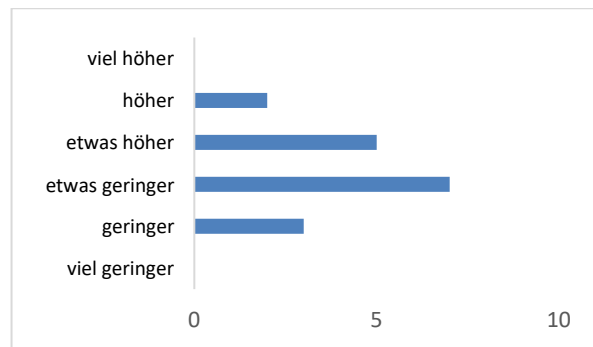


Abbildung 4: Wie ist die Qualität meiner eigenen Beiträge im Vergleich zum klassischen Review?
N = 17

Spielt Erfahrung eine Rolle?

Ein weiteres Thema ist die Altersheterogenität der PC-Mitglieder im Verhältnis zur Qualität der gutachterlichen Stellungnahmen. In den Interviews wurden unterschiedliche Meinungen geäußert, und eine befragte Person gab beispielsweise an, dass nur erfahrene Wissenschaftler*innen gutachterliche Stellungnahmen erstellen sollten. Einige andere Befragte waren der Ansicht, dass jüngere Wissenschaftler*innen etwas unsicherer seien als erfahrene. Aber die vorherrschende Meinung ist, dass Zahl und Qualität der Begutachtungsbeiträge nicht mit dem Alter zusammenhängen, denn in allen Altersgruppen gibt es mehr und weniger aktive Fachgutachter*innen, lange und kurze Beiträge sowie detaillierte und allgemeine Beiträge. Die Protokolldateien stützen dies: Wir haben die Korrelation zwischen dem Alter der Fachgutachter*innen und der Aktivitätsdauer auf der Online-Plattform berechnet, und die Korrelation betrug etwa null. Die Korrelation zwischen dem Alter der Fachgutachter*innen und der Zahl der Kommentare war gering und nicht statistisch signifikant ($r = 0,19$; $p = 0,34$).

5.3 Zahl der Fachgutachter*innen

Ein wesentliches Merkmal des PC ist, dass eine viel größere Gruppe von Peers zum Review der Anträge beiträgt als nur die ein oder zwei Fachgutachter*innen im klassischen Verfahren. Dies wurde beinahe von allen Befragten als ein großer Vorteil angesehen. Folgende Vorteile wurden genannt:

- Der Begutachtungsprozess wird so transparenter, objektiver und weniger verzerrt, was zu besseren Entscheidungen führen und positiv auf die Akzeptanz der Ergebnisse wirken kann.

- Die Beurteilung ist so weniger von der Auswahl der Fachgutachter*innen abhängig.
- Es steht mehr Fachkompetenz für die Beurteilung zur Verfügung. Zum Beispiel stammen die Antragsteller*innen in den Förderprogrammen der AvH aus zahlreichen Ländern und die Beurteilung ihrer Qualität und Karriereaussichten (anhand der Lebensläufe) erfordert Wissen über die Herkunftsländer der Antragsteller*innen. Dieses Wissen findet sich in einem PC aus 10 Personen eher als bei nur zwei Fachgutachter*innen.
- Die größere Zahl an Fachgutachter*innen führt auch dazu, dass die Forschungsthemen der Anträge ausreichend abgedeckt sind.
- Es entsteht ein Selbstkorrekturmechanismus: geringere Wahrscheinlichkeit, dass falsche Beurteilungen unbemerkt bleiben.
- Einige Ausschussmitglieder gaben an, dass durch Kommentare von einer größeren Zahl an Fachgutachter*innen das Vertrauen in das Ergebnis des Review-Verfahrens gesteigert werde.

5.4 *Benotung*

Mehrere Befragte gaben an, bei klassischen Peer Reviews seien die Noten tendenziell gut, auch wenn in den Texten der gutachterlichen Stellungnahmen Defizite angeführt würden, da die Fachgutachter*innen in der Regel die Antragsteller*innen unterstützen möchten. Die Arbeitsweise des Peer Circle fördert eine *vergleichende* Benotung, was zu stärker differenzierten und durchschnittlich schlechteren Noten führe. Dies wäre eine potenzielle Verbesserung im Entscheidungsfindungsprozess, aber wie in Abschnitt 8.2 erörtert, fehlt es dem Peer Circle derzeit an einem systematischen Benotungsverfahren.

6. Entscheidungsfindung und Auswahlergebnisse

Im vorherigen Kapitel stand die Qualität der im Peer Circle erstellten Fachgutachten im Mittelpunkt. Schwerpunkt dieses Kapitels ist die Phase nach dem PC, in der die Fachvertreter*innen dem Ausschuss einen Vorschlag unterbreiten, worauf die Diskussion im Ausschuss sowie die Entscheidung folgen.

6.1 Vorbereitung der Entscheidungsfindung

Die Rolle der Ausschussmitglieder unterschied sich von PC zu PC und von Person zu Person. Drei von ihnen begannen mit ihrer Tätigkeit, nachdem die Peer Circle die Begutachtung und Kommentierung abgeschlossen hatten, und beschränkten sich auf die klassischen Ausschussaufgaben. Die übrigen fünf fungierten auch als Fachgutachter*innen, wobei einige von ihnen sogar betonten, *alle* Anträge gelesen und begutachtet zu haben. Dies ist mit dem klassischen Peer Review-Ansatz vergleichbar, bei dem es sich häufig als unmöglich erwies, zwei externe Fachgutachter*innen zu finden. In manchen dieser Fälle wurde das Ausschussmitglied – bei fachlicher Nähe des Antrags zu seiner Fachkompetenz – zum Zweitgutachter bzw. zur Zweitgutachterin berufen.⁸

Die meisten Ausschussmitglieder schilderten in den Interviews, dass sie sich auf die Ergebnisse des Peer Circle verlassen. Viele der Argumente beziehen sich darauf, dass die Anträge im Peer Circle von mehr Fachgutachter*innen geprüft werden:

- Im PC ist es einfacher, die Argumente der Fachgutachter*innen nachzuvollziehen (als im klassischen Peer Review-Verfahren) und dann eine Rangfolge der Vorschläge zu erstellen.
- Ist ein Kommentar unklar, kann man die PC-Mitglieder um Klarstellung bitten, was beim klassischen Modell nicht so einfach ist.
- Durch die Kommentare im PC ist es einfach, die kritischen Punkte in den Anträgen zu finden.
- Wenn Konsens über die Förderung oder Ablehnung besteht, kann dieses Urteil akzeptiert werden. Dann kann man sich nur auf die Fälle mit widersprüchlichen Einschätzungen im Peer Circle und die Grenzfälle zwischen Ablehnung und Förderung konzentrieren.
- Das Problem ist weiterhin der Graubereich, in dem keine objektive Auswahl möglich ist. Dann funktioniert der PC besser, weil mehr Personen an der Begutachtung beteiligt sind.
- Mehr Sichtweisen führen zu mehr Sicherheit bei den Ausschussmitgliedern: „Wenn ich mir in einem Punkt unsicher bin, sind die PC-Kommentare sehr hilfreich bei meiner Meinungsbildung“ und „Die PC-Begutachtungen führen dazu, dass ich bei meiner eigenen Rangfolge sicherer bin.“
- Beim klassischen Ansatz besteht das Problem, dass Fachgutachten zu spät für die Entscheidungsfindung eingereicht und die Anträge dann in die nächste Auswahlrunde verschoben werden. Der Peer Circle behebt dieses Problem, da er sich auf eine Gruppentätigkeit stützt, die auch dann Ergebnisse liefert, wenn eines der Mitglieder dies nicht tut.
- Ein oder zwei Ausschussmitglieder vertreten andere Ansichten dazu, welchen Beitrag der Begutachtungsprozess im Peer Circle zu ihren Aufgaben leistet. Ihre Argumente lauten:

⁸ Wenn laut Ausschussmitglied die Argumente bei einem einzigen vorliegenden Fachgutachten klar und umfassend genug sind, um über den Antrag zu entscheiden, wird auch dann die übliche Stellungnahme für die Ausschusssitzung erstellt.

- Die Begutachtungen und Kommentare von fachfernen Mitgliedern liefern kaum einen Mehrwert.
- Die Übertragung der PC-Kommentare in einen Vorschlag ist mehr Arbeit (zwei Befragte), sie können aber zumindest zu etwas besseren Entscheidungen führen (eine befragte Person).
- Die Anträge und die vorgeschlagene Rangfolge müssen in einem interdisziplinären Ausschuss präsentiert werden, wozu andere Arten von Argumenten erforderlich sind als die Argumente, die im Peer Circle entwickelt werden.

Eines der Ausschussmitglieder schlug vor, dass ein Ausschussmitglied seine finale Rangfolge den Peer Circle-Mitgliedern vorstellen und Feedback einholen sollte. Auf diese Weise wäre der PC direkter an der Übertragung der Begutachtungsergebnisse in den Vorschlag für den Entscheidungsfindungsprozess beteiligt. Ein anderes Ausschussmitglied war der gegenteiligen Ansicht und hob Beispiele hervor, bei denen seine Einschätzung dem PC-Konsens widerspricht. Das Mitglied betonte seine Unabhängigkeit vom Peer Circle und beschrieb die Ausschussmitglieder als die „Gutachter*innen der Gutachter*innen“.

Einige der befragten Ausschussmitglieder reflektierten die Endergebnisse und gaben an, dass das Ergebnis nicht anders gewesen wäre, wenn die Anträge mit der klassischen Methode begutachtet worden wären. Aber hat sich die andere Form der Vorbereitung auf die Entscheidungsfindung ausgewirkt? Führt der Peer Circle zu mehr Einigkeit und größerer Sicherheit (wie von einigen Ausschussmitgliedern angeführt) und damit zu weniger Diskussionen im Ausschuss⁹ und folglich zu einem effizienteren Prozess? Bei der Beobachtung der Ausschusssitzungen wurde unter anderem die Zahl der Diskussions- teilnehmer*innen für jeden diskutierten Antrag erfasst. Rund 31 % der im klassischen Ausschuss präsentierten Anträge wurden nicht weiter diskutiert, im PC-Ausschuss waren es 37 %. Der auffallende Unterschied ist der Long Tail bei der klassischen Methode (Abb. 5).

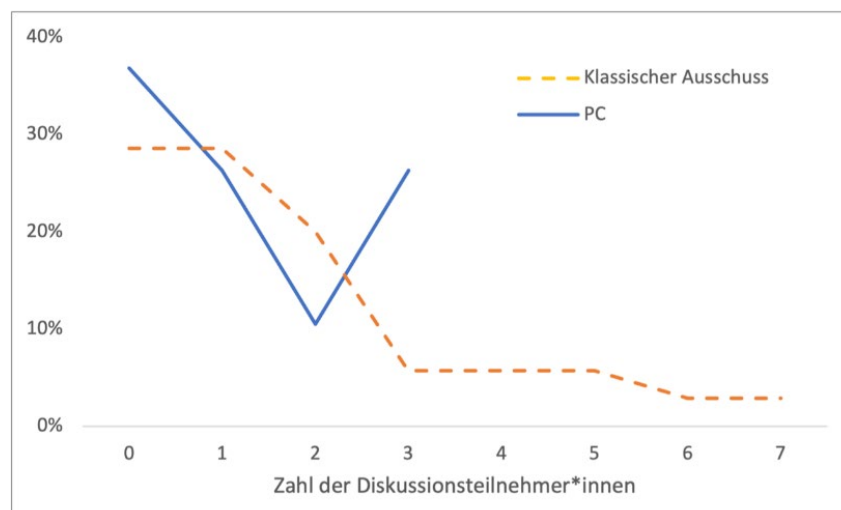


Abbildung 5: Diskussionsteilnehmer*innen pro Antrag (Sommerunde)

⁹ Frühere Untersuchungen zeigten, dass es bei Uneinigkeit in einer Gruppe, z. B. bezüglich der Qualität eines Förderantrags, mehr Diskussionen darüber gibt (Festinger, Informal social communication. In: *Psychological Review* 57 (1950) 271-280).

Die durchschnittliche Anzahl der Diskussionsteilnehmer*innen pro Antrag war im PC geringer als bei der klassischen Methode (1,26 bzw. 1,74). Die beiden Ausschüsse waren unterschiedlich groß (8 bzw. 13 Mitglieder) und bei mehr Teilnehmer*innen gibt es mehr potenzielle Diskutant*innen. Allerdings neigen Teilnehmer*innen in größeren Gruppen möglicherweise weniger dazu, einen Diskussionsbeitrag zu leisten, und nicht alle Ausschussmitglieder waren während der gesamten Sitzung anwesend. Im klassischen Ausschuss waren wesentlich mehr Anträge zu diskutieren, was zu Zeitproblemen führte. Auch dies kann die Teilnehmer*innen von einer Teilnahme an den Diskussionen abhalten. Diese Effekte können einander aufwiegen und man kann vorsichtig schlussfolgern, dass Anträge im PC-Ausschuss durchschnittlich zu weniger Diskussionen führen als in den klassischen Ausschüssen. Wir tendieren daher zu der Schlussfolgerung, dass durch das PC-Begutachtungsverfahren Unsicherheit verringert werden kann und die Ausschusssitzungen dadurch mindestens ebenso effizient sind.

6.2 Erfolgsquoten

Die durchschnittliche Erfolgsquote in den Versuchsfachgebieten und in den Kontrollfachgebieten war 2022 wesentlich geringer als 2021 (Tabelle 3). Die Tabelle zeigt auch, dass die durchschnittliche Erfolgsquote im PC 2022 etwas höher ist als in den Kontrollfachgebieten (23,6 % gegenüber 22,7 %). Die geringeren Erfolgsquoten 2022 sind zumindest teilweise auf den Rückgang des Gesamtverhältnisses von Gesamtfördervolumen zur Gesamtzahl Anträgen zurückzuführen.

Tabelle 3: Erfolgsquote in PC-Gebieten und in den klassisch begutachteten Gebieten vor und während PC

	Erfolgsquote vor PC 2021	Erfolgsquote PC-Zeitraum 2022
PC-Fachgebiete	28,9 %	23,6 %
Kontrollfachgebiete	33,3 %	22,7 %
Alle anderen Fachgebiete	33,7 %	27,6 %

N= 321

Warum ist der Rückgang in den Kontrollfachgebieten stärker als in den Versuchsfachgebieten? Für die endgültigen Entscheidungen werden die Rangfolgen aller Ausschussgruppen zusammengenommen. Auch wenn die durchschnittliche Punktzahl für ein Fachgebiet gleichbleibt, sinkt die Erfolgsquote, wenn andere Fachgebiete im Durchschnitt steigende Werte aufweisen. Tabelle 4 zeigt die durchschnittlichen Punktzahlen der Fachgebiete für die vier Bedingungen. In den PC-Fachgebieten bleibt die durchschnittliche Ausschusswertung gleich (0,54). In den Kontrollfachgebieten ging der Durchschnittswert mit 18 % (von 0,67 auf 0,55) erheblich zurück.

Tabelle 4: Durchschnittliche Ausschusswertung vor und während PC

	Ausschusswertung vor PC 2021			Ausschusswertung PC-Zeitraum 2022		
	N	Mittelwert	Std. Abw.	N	Mittelwert	Std. Abw.
PC-Fachgebiete	97	0,54	0,79	89	0,54	0,72
Kontrollfachgebiete	69	0,67	0,84	70	0,55	0,78
Alle anderen Fachgebiete	926	0,65	0,83	822	0,61	0,81

Den S-Fällen wurde ein Wert von 2 und den A-Fällen ein Wert von 0 zugewiesen.

Da die durchschnittlichen Punktzahlen der PC-Fachgebiete und aller anderen Fachgebiete etwa gleichbleiben, erklären nicht die Werte, sondern die verfügbaren Fördermittel den

Rückgang der Erfolgsquote für die PC-Fachgebiete. In den Kontrollfachgebieten ging der mittlere Wert signifikant zurück, was erklärt, dass die Erfolgsquote dieser Fachgebiete wesentlich stärker zurückging als die Erfolgsquote der Versuchsfachgebiete.

6.3 Qualität der Anträge

Zur Messung und zum Vergleich der Qualität der Anträge werden zusätzlich zu den Wertungen der Ausschusssmitglieder für die Beurteilungskriterien objektivere Daten benötigt. Aus der Kombination von Lebenslauf und bibliometrischen Daten lassen sich verschiedene faktische Daten zu Mobilität, Qualität des Netzwerks der Antragsteller*innen und Karriere ableiten, aber dies geht über den Rahmen dieser Studie hinaus. Wir beschränken uns hier auf die bisherige Forschungsleistung der Antragsteller*innen, da die Interviews und die Beobachtungen der Ausschusssdiskussionen deutlich gezeigt haben, dass Produktivität und Impact (Status) der Zeitschriften eine Rolle spielen. Es wurden bibliometrische Daten gesammelt und Leistungskennzahlen berechnet. Die Produktivität wird auf zwei Arten gemessen:

- Die Zahl der Publikationen in internationalen Zeitschriften in der Scopus-Datenbank.
- Die Zahl der Publikationen bei fraktionierter Berechnung. Bei dieser Kennzahl wird die Zahl der Koautor*innen berücksichtigt: Den einzelnen Autor*innen wird jeweils nur ein Anteil jeder Publikation zugeordnet, der durch die gleichmäßige Verteilung der Publikation auf die Autor*innen berechnet wird. Bei einer Publikation eines Antragstellers bzw. einer Antragstellerin mit 5 Koautor*innen wird nur 1/6 der Publikation dem Antragsteller bzw. der Antragstellerin zugeordnet.

Der Impact der Veröffentlichungen von Antragsteller*innen wird anhand der folgenden Kriterien gemessen:

- Anteil und Anzahl von 1 % der am häufigsten zitierten Publikationen, normalisiert nach Fachgebiet, Veröffentlichungsart und Veröffentlichungsjahr.
- Anteil und Anzahl von 5 % der am häufigsten zitierten Publikationen, ebenso normalisiert.
- Summe der SNIP-Werte. Der SNIP-Wert misst den Impact einer Zeitschrift (es wird also jeder Zeitschrift ein solcher Wert zugewiesen), und zur Berechnung des gesamten Impacts des Antragstellers bzw. der Antragstellerin werden die SNIP-Werte addiert.

Aus Zeitgründen beschränken wir die Analyse auf die beiden Chemie-Fachgebiete. Chemie ist dafür eine gute Wahl, da in diesem Bereich Zeitschriftenpublikationen die Standardkommunikationsform sind. Die erfahrenen Antragsteller*innen berücksichtigen wir nicht, da es in der Stichprobe zu wenige von ihnen gibt. Auch die Zahl der verbleibenden Postdoc-Antragsteller*innen ist nicht hoch und die Ergebnisse sind entsprechend mit Vorsicht zu interpretieren.¹⁰ Tabelle 5 zeigt die Ergebnisse für die beiden Chemie-Felder 2021 und 2022.

¹⁰ Wir haben hier die geförderten und die nicht geförderten Antragsteller*innen in diesen beiden Fachgebieten 2021 und 2022 aufgenommen. Die Zahlen lassen sich daher nicht einfach mit den bibliometrischen Daten aus der unlängst durchgeführten Evaluation des HFST-Programms vergleichen, da in dieser Studie nur geförderte Antragsteller*innen berücksichtigt werden (Geyer, Grasenick, Kleinberger-Pierer, Gorriaz (2021), *Evaluation des Humboldt-Forschungsstipendien-Programms (HFST) der Alexander von Humboldt- Stiftung*. Endbericht. Wien: Inspire).

Tabelle 5: Durchschnittswerte der bibliometrischen Leistungskennzahlen, zwei Chemie-Fachgebiete*

	Anorganische Chemie 2021 (N=28)			Festkörperchemie 2021 (N=18)		
	Mittelwert	Median	Std. Abw.	Mittelwert	Median	Std. Abw.
Alle Pub.	19,9	13,5	29,7	16,6	17,5	6,4
Fraktioniert	3,5	2,8	4,1	2,8	2,6	1,4
Snip-Summe	28,5	15,8	50,1	29,4	26,8	18,9
Anteil beste 1 %	2,9 %	0,0 %	0,1	6,0 %	0,0 %	0,1
Anteil beste 5 %	14,2 %	14,0 %	0,1	20,6 %	11,8 %	0,2 %
	Anorganische Chemie 2022 (N=17)			Festkörperchemie 2022 (N=9)		
	Mittelwert	Median	Std. Abw.	Mittelwert	Median	Std. Abw.
Alle Pub.	18,4	9	18,6	18,67	14	13,5
Fraktioniert	3,3	1,8	3,8	3,6	2,6	2,1
Snip-Summe	22,1	12,3	27,2	18,5	13,9	13,1
Anteil beste 1 %	1,9 %	0,0 %	0,1	1,7 %	0,0 %	0,0
Anteil beste 5 %	11,3 %	7,7 %	0,1	16,0 %	10,0 %	0,2 %

* Nur die Postdoktorand*innen

Die Tabelle zeigt, dass die Antragsteller*innen im Durchschnitt recht gute Werte erzielen, was sich bei den Werten für den Anteil an den am häufigsten zitierten 1 % und 5 % der Publikationen besonders deutlich erkennen lässt. Die Werte sind deutlich höher als erwartet (erwartet wurden selbstverständlich 1 % bzw. 5 %). Beispielsweise fallen 2021 im Bereich Anorganische Chemie (vor dem PC) im Durchschnitt 2,9 % der Publikationen in die Klasse der 1 % meistzitierten Publikationen, 14,2 % in die der besten 5 % und damit jeweils dreimal so viel wie erwartet. Die Durchschnittswerte der meisten Leistungskennzahlen für Anorganische Chemie sind 2022 etwas niedriger als 2021, und dies gilt auch (für die Impact-Kennzahlen) für Festkörperchemie.

Die Ausschusswertungen weisen ein ähnliches Muster auf: Der Durchschnittswert für Anorganische Chemie ging von 0,59 auf 0,45 zurück, was auf eine insgesamt niedrigere Qualität der Anträge im Jahr 2022 hindeutet. Der Durchschnittswert für Festkörperchemie sinkt in etwa gleich stark (von 0,84 auf 0,63) und die Standardabweichungen zeigen für beide Fachgebiete eine recht große Variation. Die geringere durchschnittliche bibliometrische Leistung spiegelt sich in den niedrigeren durchschnittlichen Ausschusswertungen wider; die Ergebnisse zeigen also, dass die Ausschusswertungen in nicht unerheblichem Maße durch objektive Kennzahlen gestützt werden.

Tabelle 6: Durchschnittliche Ausschusswertungen für Anorganische Chemie und Festkörperchemie

	Vor PC 2021		Nach PC 2022	
	Mittlere Ausschusswertung		Mittlere Ausschusswertung	
	Mittelwert	Standardabweichung	Mittelwert	Standardabweichung
Anorganische Chemie	0,59	0,80	0,45	0,69
Festkörperchemie	0,84	0,91	0,63	0,95

Den S-Fällen wurde ein Wert von 2 und den A-Fällen ein Wert von 0 zugewiesen.

N = 72. Nur Postdoktorand*innen

Die gutachterlichen Stellungnahmen und die Ausschusssitzungen lassen annehmen, dass der Publikationsoutput ein wichtiger Faktor bei der Beurteilung der Anträge ist. Begriffe wie „Länge der Publikationsliste“, „beste Publikationen“ und „beste Zeitschriften“ fallen oft, und zwar häufiger als Begriffe zu beinahe allen anderen Beurteilungsdimensionen (Tabelle 2).

Zwar werden im Begutachtungsprozess keine bibliometrischen Kennzahlen verwendet, aber es wäre zu erwarten, dass geförderte Antragsteller*innen bei diesen Kennzahlen im Durchschnitt höhere Werte erzielen als die abgelehnten Antragsteller*innen.¹¹ Der Begriff „Durchschnitt“ ist wichtig, da die Verwendung dieser Kennzahlen auf individueller Ebene umstritten ist, sie aber auf Gruppenebene sinnvoll sind.¹²

Tabelle 7: Vergleich der bibliometrischen Werte für geförderte und abgelehnte Antragsteller*innen, zwei Chemie-Fachgebiete

		2021		2022	
		Median	Durchschnitt	Median	Durchschnitt
Anorganische Chemie	# Publikationen	Gefördert*	Abgelehnt	Gleich	Abgelehnt
	# Anteilige Publikat.&	Gleich	Abgelehnt	Gefördert	Abgelehnt
	# Snip-Summe	Gefördert	Abgelehnt	Gefördert	Abgelehnt
	# meistzitierte 5 %	Gefördert	Abgelehnt	Gefördert	Gefördert
	# meistzitierte 1 %	Gleich	Abgelehnt	Gleich	Gefördert
	# Gefördert	3**		5	
	# Neutral	2		2	
	# Abgelehnt	5		3	
	Festkörperchemie	# Publikationen	Gefördert	Gefördert	Gefördert
# Anteilige Publikat.&		Gefördert	Gefördert	Abgelehnt	Abgelehnt
# Snip-Summe		Gefördert	Gefördert	Gefördert	Gefördert
# meistzitierte 5 %		Gefördert	Gefördert	Gefördert	Abgelehnt
# meistzitierte 1 %		Gefördert	Gefördert	Gleich	Gefördert
# Gefördert		10		5	
# Abgelehnt		0		3	

N = 72. Nur die Postdoktorand*innen unter den Antragsteller*innen

N pro Gruppe (IC = Anorganische Chemie; SC = Festkörperchemie)

IC 2021 gefördert = 9; IC 2021 abgelehnt = 19; IC 2022 gefördert = 3; IC 2022 abgelehnt = 14;

SC 2021 gefördert = 8; SC 2021 abgelehnt = 10; SC 2022 gefördert = 3; SC 2024 abgelehnt = 6.

* Diese Zellen geben an, welche Gruppe einen höheren Wert erreicht. Gleich bedeutet circa der gleiche Wert.

** Im Bereich Anorganische Chemie erzielten die geförderten Antragsteller*innen 2021 bei drei Kennzahlen bessere Werte als die abgelehnten Antragsteller*innen, fünf Mal weniger gute und zwei Mal etwa gleiche.

& Publikationen als fraktionierte Berechnung = bei einer Publikation mit fünf Autor*innen werden dem Autor / der Autorin 20 % dieser Veröffentlichung zugeordnet

Weisen diejenigen, die ein Stipendium erhalten, im Durchschnitt bessere Leistungen auf als diejenigen, die nicht erfolgreich sind? Tabelle 7 zeigt, in welchen Bereichen geförderte Antragsteller*innen bessere Werte erzielen als die nicht geförderten, wo es umgekehrt ist und wo beide Gruppen etwa gleichauf liegen. Die Tabellen zeigen ein gemischtes Bild, was auch wieder an den geringen Zahlen liegen kann. Im Bereich Anorganische Chemie 2021 haben die nicht Geförderten bei fünf von zehn Kennzahlen einen besseren Wert, die Geförderten haben bei drei Kennzahlen einen besseren Wert und bei zwei Kennzahlen haben beide Gruppen ungefähr denselben Wert. Im Peer Circle (Anorganische Chemie 2022) erzielten die

¹¹ Da wir hier alle abgelehnten Antragssteller*innen berücksichtigen, kann der Durchschnitt der Gruppe der abgelehnten Antragsteller*innen aufgrund einiger Antragsteller*innen mit sehr niedrigen Werten gering sein. Würde man die Analyse auf die abgelehnten Antragsteller*innen mit besseren Leistungen beschränken, sähe das Bild möglicherweise anders aus. Siehe z. B. Peter Van den Besselaar, Loet Leydesdorff, Past performance, peer review, and project selection, *Research Evaluation* 18 (2009) issue 4, 273-288.

¹² Leistungskennzahlen weisen auch deutliche Unterschiede für Antragsteller*innen aus unterschiedlichen Ländern/Regionen auf. Aufgrund der geringen Zahlen können wir diese Differenzierung hier nicht aufnehmen.

Stipendiat*innen bei fünf Kennzahlen bessere Werte, die Abgelehnten bei drei, und wiederum sind bei zwei Kennzahlen die Werte etwa gleich. Interessanterweise erzielen im PC (2022) die Stipendiat*innen geringere Werte bei den Publikationszahlen und beim Zeitschriften-Impact, aber höhere Werte bei der Anzahl der meistzitierten Publikationen. Die Ergebnisse für Festkörperchemie sind etwas anders: 2021 haben die geförderten Antragsteller*innen bei allen zehn Kennzahlen bessere Werte, aber 2022 zeigt sich das gleiche Bild wie für Anorganische Chemie. Diese Ergebnisse deuten darauf hin, dass sich die PC-Ergebnisse nicht von den anderen Ergebnissen unterscheiden.

Zusammengefasst: (i) Die Gesamtqualität der Antragsteller*innen im Bereich Chemie ist hoch, 2022 jedoch etwas geringer als 2021; (ii) die durchschnittliche Punktzahl der Antragsteller*innen seitens der Ausschüsse spiegelt die durchschnittlichen bibliometrischen Werte wider, (iii) die bibliometrischen Werte der geförderten Antragsteller*innen sind nicht insgesamt höher als die der nicht geförderten Antragsteller*innen, was unerwartet ist. Dies deutet darauf hin, dass Produktivität, Impact und Qualität der Zeitschriften im Entscheidungsfindungsprozess weniger wichtig sind als die Gutachten und die Ausschusssitzungen vermuten lassen. Es wäre gut, diese Analyse mit mehr Peer Circle- und Kontrollgruppen zu wiederholen, um die Einschränkungen aufgrund der geringen Fallzahlen zu überwinden.

6.4 Vielfalt der ausgewählten Anträge

Eine wichtige Frage ist, ob sich die Organisationsform des Auswahlprozesses auf das Ergebnis in Bezug auf (Un)Gleichstellung der Geschlechter und geschlechtsbezogene Verzerrungseffekte auswirkt. Ungleichstellung der Geschlechter ist als unterschiedliche Erfolgsquote definiert, während eine Verzerrung auftritt, wenn dieser Unterschied nicht auf Verdiensten basiert. Wir stellen die Erfolgsquoten nach Geschlecht dar, können aber nicht prüfen, ob sie auf Unterschieden bei der bisherigen Leistung beruhen. Dies liegt daran, dass die Fallzahlen zu klein werden: Im Peer Circle Anorganische Chemie (2022) gibt es nur sechs Frauen; in der Gruppe Anorganische Chemie 2021 und in den Gruppen Festkörperchemie 2022 und 2021 gibt es acht, acht bzw. drei Frauen.

Insgesamt ist bei Männern und Frauen die Wahrscheinlichkeit, eine HFST-Förderung zu erhalten, gleich hoch und bei beiden Geschlechtern beträgt die Gesamterfolgsquote in unserer Stichprobe von 321 Antragsteller*innen 27 %. Aufgeschlüsselt nach Jahr und Begutachtungsmodell stellen wir fest, dass in den PC-Fachgebieten 2021 die Erfolgsquote von Männern wesentlich höher, 2022 hingegen die Erfolgsquote von Frauen leicht höher ist. In den Kontrollfachgebieten ist 2021 die Erfolgsquote von Frauen wesentlich höher und 2022 fällt der Unterschied bei den Erfolgsquoten geringer aus (Tabelle 8).

Tabelle 8: Erfolgsquoten von Männern und Frauen 2021 und 2022, zwei Chemie-Fachgebiete

		Erfolgsquote vor PC 2021	Erfolgsquote nach PC 2022
PC-Fachgebiete	Frauen	20,0 %	24,1 %
	Männer	32,8 %	23,3 %
	Summe	28,9 %	23,6 %
Kontrollfachgebiete	Frauen	41,7 %	25,0 %
	Männer	28,9 %	21,4 %
	Summe	33,3 %	22,7 %

N=321

Hierbei handelt es sich um aggregierte Erfolgsquoten und ihnen liegen sehr unterschiedliche Erfolgsquoten auf Fachgebietsebene zugrunde. Tabelle 9 – nur mit den Postdoktorand*innen – zeigt, dass es große Schwankungen gibt, was auch daran liegt, dass schon eine erfolgreiche Frau mehr oder weniger das Ungleichgewicht in die andere Richtung verschieben kann. In sechs von acht Fachgebieten verändert sich das Ungleichgewicht von einem Jahr zum anderen: In einem Jahr sind die Frauen unter den Stipendiat*innen unterrepräsentiert und im anderen Jahr gleichermaßen vertreten oder überrepräsentiert. Nur in den Bereichen Neuere Geschichte (Frauen unterrepräsentiert) und Werkstofftechnik (Frauen überrepräsentiert) hat sich das Ungleichgewicht der Geschlechter nicht verändert. Es wäre interessant, die Geschlechterunterschiede auf Fachgebietsebene mit den Leistungsunterschieden zu vergleichen, da dies mehr über (die Abwesenheit von) Verzerrungen aussagen würde. Aber dies geht über den Rahmen dieses Berichts hinaus.

Tabelle 9: Erfolgsquote nach Geschlecht, Fachgebiet und Jahr*

	2021		2022		Summe	
	N	Anteil	N	Anteil	N	Anteil
Frauen						
Anorganische Chemie	8	12,5 %	6	33,3 %	14	21,4 %
Materialwissenschaft	2	50,0 %	4	25,0 %	6	33,3 %
Neuere Geschichte	3	0,0 %	2	0,0 %	5	0,0 %
Zoologie	11	18,2 %	10	20,0 %	21	19,0 %
Alte Geschichte	4	50,0 %	3	33,3 %	7	42,9 %
Pflanzenwissenschaft	7	57,1 %	6	16,7 %	13	38,5 %
Festkörperchemie	8	37,5 %	3	33,3 %	11	36,4 %
Werkstofftechnik	3	33,3 %	4	25,0 %	7	28,6 %
Männer						
Anorganische Chemie	20	40,0 %	11	9,1 %	31	29,0 %
Materialwissenschaft	9	33,3 %	13	30,8 %	22	31,8 %
Neuere Geschichte	16	31,3 %	8	37,5 %	24	33,3 %
Zoologie	11	45,5 %	14	21,4 %	25	32,0 %
Alte Geschichte	10	20,0 %	5	40,0 %	15	26,7 %
Pflanzenwissenschaft	7	14,3 %	7	28,6 %	14	21,4 %
Festkörperchemie	10	50,0 %	6	33,3 %	16	43,8 %
Werkstofftechnik	10	20,0 %	11	0,0 %	21	9,5 %

* Nur Postdoktorand*innen.

Rot/blau: Frauen weisen eine niedrigere/höhere Erfolgsquote als Männer auf.

Diese Ergebnisse weisen zusammengefasst darauf hin, dass das Geschlechtergleichgewicht auf Fachgebietsebene fluktuiert, dass aber diese Veränderungen auf höheren Aggregations-ebenen verschwinden und sich ein insgesamt ausgewogenes Bild ergibt. Ein Vorbehalt sind die geringen Zahlen, weshalb die festgestellten Ungleichheiten ebenso gut auf zufällige Abweichungen wie auf geschlechtsbezogene Verzerrungseffekte zurückzuführen sein können. Wie so häufig wären mehr Daten die Lösung. Es lassen sich keine spezifischen Schlussfolgerungen zum Peer Circle ziehen.

7. Aktivität der Fachgutachter*innen

7.1 Unterschiede zwischen den Fachgebieten?

Da die Aktivitäten und Arbeitsweisen der Fachgutachter*innen vom Fachgebiet abhängen können, behandeln wir zunächst die Frage, ob es bei der Einschätzung des PC Unterschiede zwischen den Fachgebieten gibt. Bei den Durchschnittswerten der verschiedenen Erhebungselemente gibt es keine großen Unterschiede zwischen den vier Versuchsfachgebieten. Nur bei wenigen Elementen beträgt der Unterschied zwischen dem höchsten und dem niedrigsten Wert mehr als 40 % und in allen außer einem dieser Fälle weist die *Materialwissenschaft* einen wesentlich niedrigeren Wert auf als die anderen Fachgebiete. Alle diese Elemente beziehen sich auf die Arbeitsweise des PC (s. Abschnitt 8.1) und die Unterschiede zwischen den Fachgebieten bei den stärker bewertenden Erhebungselementen zur Qualität des Peer Circle sind gering. Die Zahl der PC-Mitglieder ist begrenzt und die Ergebnisse sollten nicht über die Stichprobe hinaus verallgemeinert werden.

7.2 Peer Circle-Aktivitäten im Zeitverlauf

Wir haben die Protokolldateien der Online-Plattform analysiert, um das Muster und die Intensität der Aktivität im PC darzustellen. Tabelle 10 zeigt die Intensität der Begutachtungsaktivitäten in jedem Peer Circle in der ersten Evaluationsrunde (Sommer 2022) und der zweiten Evaluationsrunde (Herbst 2022). In der Sommerrunde gaben die Fachgutachter*innen im Durchschnitt 35 Kommentare ab, die Ausschussmitglieder 25 und die AvH-Beschäftigten rund 45¹³. Es gibt erhebliche Unterschiede zwischen den Teilnehmer*innen: von nur wenigen Kommentaren bis hin zu mehr als 70.

Tabelle 10: Anzahl der Kommentare nach Fachgebiet und Rolle

	Durchschnitt		pro Antrag		Minimum		Maximum	
	Sommer	Herbst	Sommer	Herbst	Sommer	Herbst	Sommer	Herbst
Anorganische Chemie – Fachgutachter*innen	37	30	3,7	3,0	11	6	59	47
Anorganische Chemie – Ausschussmitglieder	43	23	4,3	2,3	15	13	70	33
Anorganische Chemie – AvH-Beschäftigte	36	37	3,6	3,7	2	5	70	68
Materialwissenschaft – Fachgutachter*innen	31	20	3,1	2,2	15	4	51	37
Materialwissenschaft – Ausschussmitglieder	16	5	1,6	0,6	16	3	16	6
Materialwissenschaft – AvH-Beschäftigte	41	40	4,1	4,4	3	7	79	72
Zoologie – Fachgutachter*innen	36	31	2,3	3,4	15	12	65	52
Zoologie – Ausschussmitglieder	37	79	2,3	8,8	37	79	37	79
Zoologie – AvH-Beschäftigte	59	49	3,7	5,4	29	10	88	88
Neuere Geschichte – Fachgutachter*innen	38	22	2,7	1,6	8	11	71	36
Neuere Geschichte – Ausschussmitglieder	11	5	0,8	0,4	7	5	14	5
Neuere Geschichte – AvH-Beschäftigte	44	31	3,1	2,2	26	1	61	60

Quelle: Zusammenstellung aus den Plattform-Protokolldateien.

Ein weiterer Unterschied ist die Involvierung der Ausschussmitglieder. Beim klassischen Verfahren können die Ausschussmitglieder auch als Fachgutachter*innen fungieren, wenn ihr Fachgebiet dem Thema des Antrags sehr nahekommt. Im PC ist dies auch der Fall, und

¹³ Hierzu zählen auch koordinierende (nicht inhaltsbezogene) Kommentare, wie die Aufforderung an die Fachgutachter*innen, bestimmte Themen anzusprechen.

einige, aber nicht alle Ausschussmitglieder sind im Begutachtungsprozess aktiv. In der Zoologie waren die Ausschussmitglieder sehr aktiv (zusammen 6,2 Stunden in Runde 1 und 7,8 Stunden in Runde 2). An nächster Stelle kommt Anorganische Chemie (5,9 h; 4,3 h), gefolgt von Materialwissenschaft (2,5 h; 4,5 h) und Geschichte (0,9 h; 1,2 h).

Die Veränderung sieht folgendermaßen aus: In allen PC ging die Anzahl der Kommentare bei den Fachgutachter*innen und den Ausschussmitgliedern zurück, nur im Fachgebiet Zoologie nahm sie zu. Bei den AvH-Beschäftigten sieht das Bild gemischt aus. Der Gesamtrückgang ist, wie die Spalte „pro Antrag“ (Tabelle 10) zeigt, nicht auf eine Abnahme der Anzahl der Anträge zurückzuführen.

Die Abnahme der Anzahl der Kommentare pro Antrag wurde nicht durch längere Kommentare ausgeglichen: Die durchschnittliche Länge der Kommentare ging von Runde 1 zu Runde 2 von 783 auf 712 Wörter leicht zurück. Wie wir unten sehen, ging auch die für die Begutachtung der Anträge aufgewandte Zeit in Runde 2 gegenüber Runde 1 zurück. Die Fachgutachter*innen führten dies zumindest teilweise auf die Planung der zweiten Runde zurück: Die erste Hälfte der zweiten Runde begann Mitte August und lag damit im Sommerurlaub (August) vieler PC-Mitglieder, in dem Zeitraum, in dem viele wissenschaftliche Konferenzen stattfinden (September) und im Beginn der Vorlesungszeit.

Bei der zeitlichen Verteilung der PC-Aktivitäten in der ersten Runde gab es gewisse Unterschiede zwischen den Peer Circle (Übersicht 1). In der Anorganischen Chemie begannen einige Fachgutachter*innen früh, aber dann gab es einen langen Zeitraum von sehr geringer Aktivität, ehe der PC am Ende des Zeitraums aktiv wurde. Gleiches gilt für die Materialwissenschaft. Im Fachgebiet Neuere Geschichte konzentrierten sich die Aktivitäten auf den Anfang und das Ende, und im Zoologie-PC konzentrierten sich die Aktivitäten ebenfalls auf den Anfang, waren aber anschließend auf die zweite Hälfte des verfügbaren Zeitraums verteilt. Wir haben die gleiche Analyse für die Herbstrunde durchgeführt. Das Muster sieht etwas anders aus: Im Fachgebiet Zoologie verteilten sich die Aktivitäten auf einen langen Zeitraum, mit einer leichten Konzentration in der Mitte, und in der Materialwissenschaft gab es zu Beginn keinerlei Aktivität, sondern alle Aktivitäten fanden in der zweiten Hälfte des Begutachtungszeitraums statt. Insgesamt konzentrierten sich die Aktivitäten in einigen, aber nicht in allen Gremien im oder um den interaktiven Zeitraum. Dies traf auf den zweiten Begutachtungszeitraum weniger stark zu als auf den ersten.

Legende: Übersicht 1 und 2 – auf den nächsten zwei Seiten

Tabellen auf den nächsten zwei Seiten (basierend auf den Plattform-Protokolldateien):

- Die oberste Zeile gibt die Wochen an: Woche 1, Woche 2 usw.
- Dann die drei Gruppen pro Peer Circle:
 - o die weißen Linien sind die Fachgutachter*innen,
 - o die grünen Linien sind die Ausschussmitglieder,
 - o die gelben Linien sind die AvH-Beschäftigten.
- Dies wird für jedes der vier Felder wiederholt.
- Die Zahl in einer Zelle zeigt für das betreffende PC-Mitglied an, wie viele Stellungnahmen oder Kommentare in dieser Woche abgegeben wurden.
- Die roten Ovale stehen für Zeiträume mit intensiver Aktivität.
- Die grauen Streifen markieren „interaktive Zeiträume“.

PC	role	04,06,07,08,09,10,11,12,13,14,15,16,17,18,21,22,23,24,25,27,28,29,30,31,01,02,03,04,07,08,11,12,14,17,19,20,21,22,23,24,25,26,27,28,30,01,02,03,05,06,08,09,10,11,12,13,14,15,16,17,18,19,20,23,24,25,30,01,04,06,07,09,10,13,14,17,19,20,21,Gr
chem	FG	1 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
chem	FG	1
chem	FG	2 38
chem	FG	2
chem	FG	1
chem	FG	2
chem	FG	1
chem	FG	1
chem	2FV	5 1 1 1 1 1 3 1
chem	2AVH	4 2 10 3
material	FG	1
material	FG	4 10 3
material	FG	1
material	FG	1
material	FG	1
material	FG	1
material	FG	1
material	FG	1
material	FG	1
material	2FV	2
material	2AVH	1 8 3 1
zoool	FG	1 5 3 6 13 10 1 1 15
zoool	FG	4 1 1 1
zoool	FG	4
zoool	FG	5
zoool	FG	13 5 2 5
zoool	2FV	4 1
zoool	2AVH	12 1 5 1 7 1 1 1 5 1
hist	FG	11 4 9 21 14
hist	FG	1 5 1 4 2 2
hist	FG	14 4 7
hist	FG	2 8 2 2 4 4 2
hist	FG	1 8 4 12 3 2 7 3 6
hist	FG	4 4
hist	2FV	9 15 2 8 2 2 1
hist	2AVH	10 1 77 80 38 45 22 41 11 20 60 19 28 34 16 18 3 7 4 2 1 2 9 10 3 4 1 3 5 92 6 6 29 6 15 3 29 6 2 42 7 14 2 1 5 4 11 2 7 25 11 75 87 35 72 45 49 6 96 73 76 50 14 20 3 3 1 54 7 5 5 10 5 4 1 2 1 15 6

Übersicht 1: Aktivität im Zeitverlauf – Sommerrunde

7.3 Vorgehensweise und Integration der Peer Circle-Aktivitäten

Ein wichtiger Aspekt des PC-Konzepts besteht darin, dass es sich um einen gemeinsamen Begutachtungsprozess handelt, bei dem nicht alle PC-Mitglieder alle Teile eines Antrags begutachten müssen. Der Erstgutachter bzw. die Erstgutachterin (in der Regel die Person, deren Fachgebiet dem Thema des Antrags am nächsten kommt) soll beginnen und alle Begutachtungsfragen beantworten, und die anderen sollen die Begutachtungsfragen beantworten, insoweit sie sich dazu für qualifiziert erachten, und die Stellungnahmen anderer Mitwirkender kommentieren. Wie bereits beschrieben, haben die Erstgutachter*innen nicht immer früh im Prozess angefangen, was die Arbeit der anderen PC-Mitglieder beeinflusst hat. Mehr als die Hälfte der Fachgutachter*innen fingen erst oder tendenziell erst an, nachdem der zugeteilte Erstgutachter bzw. die zugeteilte Erstgutachterin tätig geworden war, während weniger als die Hälfte tendenziell nicht auf den*die Erstgutachter*in wartete (Abb. 6).

Ein weiterer Aspekt des PC besteht darin, dass auch diejenigen, deren Fachgebiet nicht dem Projektvorschlag entspricht, dennoch einen sinnvollen Beitrag leisten können, beispielsweise durch die Beurteilung des Lebenslaufs, aber auch von Teilen des fachlichen Inhalts, wie Methodik oder (gesellschaftliche) Relevanz. Wenn Fachgutachter*innen keine Expert*innen für das Thema des Forschungsprojekts sind, kennen sie möglicherweise das Heimatland des Antragstellers bzw. der Antragstellerin und können somit dessen/deren Karriere beurteilen. In der Tat haben sich PC-Mitglieder an der Beurteilung dieser Anträge beteiligt, indem sie den ganzen Antrag oder Teile davon gelesen und Kommentare zu einigen Begutachtungsfragen hinzugefügt haben (Abb. 7).

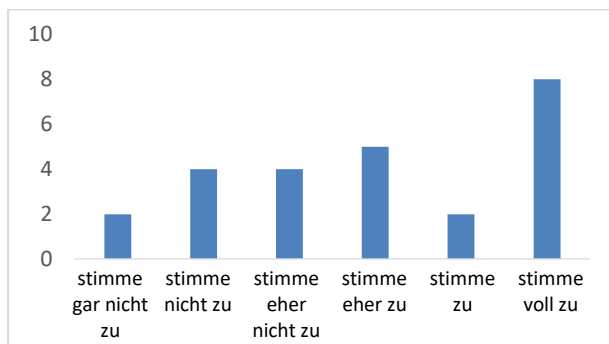


Abbildung 6: Ich fing mit meiner Begutachtung nach dem*r Erstgutachter*in an.

N = 25

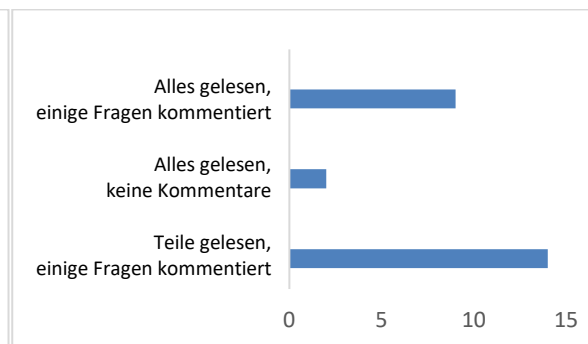


Abbildung 7: Wie gingen Sie mit Anträgen um, die weit von Ihrem Fachgebiet entfernt waren?

N = 25

Die Befragten gaben an, den Antrag vollständig gelesen zu haben, wenn ihnen die Rolle des Erstgutachters bzw. der Erstgutachterin zugewiesen wurde. Der PC-Ansatz hat es jedoch tatsächlich möglich gemacht, sich am Begutachtungsprozess zu beteiligen, ohne alle Anträge vollständig zu lesen. Etwa ein Viertel der beteiligten PC-Mitglieder hat alle Anträge gelesen (Abb. 8). In den Interviews wurde auch deutlich, dass es einfacher war, den Lebenslauf und die Zukunftsperspektiven zu beurteilen als den Projektvorschlag – da die PC-Mitglieder dazu einschlägige Fachkenntnisse benötigten.

Schließlich waren die Fachgutachter*innen zuversichtlich, beitragen zu können, auch wenn sie kein echter Peer waren. Abbildung 9 zeigt, dass drei Viertel der PC-Mitglieder etwas bis vollkommen zustimmen, dass es in diesen Fällen einfach war, einen Beitrag zur Beurteilung

zu leisten. Die Interviews deuten darauf hin, dass sich einige der weniger erfahrenen PC-Mitglieder hinsichtlich ihrer Rolle unsicherer waren als die erfahreneren Mitglieder.

Allgemein lässt sich schlussfolgern, dass sich die Arbeit entsprechend den unterschiedlichen Fachkompetenzen auf die verschiedenen PC-Mitglieder verteilte – was eines der Ziele des Peer Circle ist.

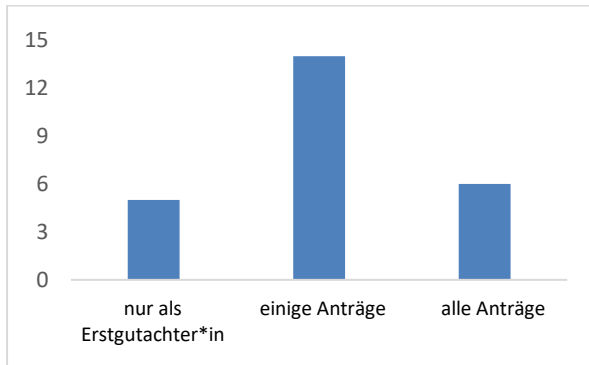


Abbildung 8: Ich habe die Anträge vollständig gelesen
N = 25

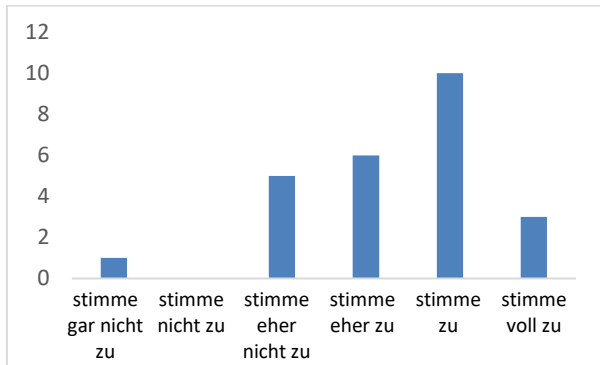


Abbildung 9: Es war einfach, als nicht wirklicher Peer einen Beitrag zu leisten, da andere ihre Meinung hinzufügen konnten
N = 25

Beim klassischen Begutachtungsverfahren tragen die Fachgutachter*innen die alleinige Verantwortung für die Qualität des Fachgutachtens, beim PC hingegen gibt es eine kollektive Verantwortung. Daher können die einzelnen PC-Mitglieder möglicherweise weniger Verantwortung für das Ergebnis empfinden. Dies trifft auf etwas mehr als die Hälfte der Fachgutachter*innen zu, auf die übrigen nicht (Abb. 10).

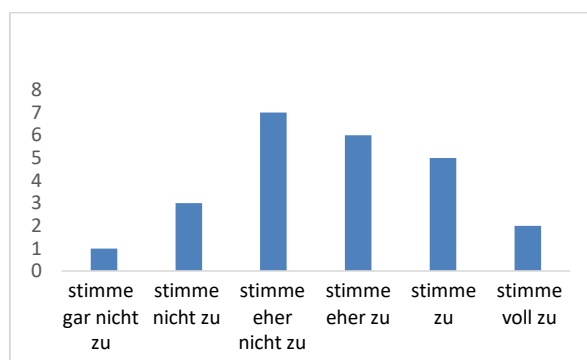


Abbildung 10: Da die Begutachtung eine Gruppentätigkeit ist, empfand ich weniger Verantwortung für die Ergebnisse
N = 24

Die Interviews ergaben, dass die PC-Mitglieder ihre Tätigkeiten sehr unterschiedlich planen. Einige beginnen sofort, andere warten auf die interaktiven Zeiträume, und dann gibt es diejenigen, die erst gegen Ende des Begutachtungszeitraums kurz vor Fristablauf anfangen, was die gegenseitige Kommentierung der Beiträge erschwert. Besonders problematisch für die Integration der PC-Tätigkeiten war es, wenn einer der Letzteren als Erstgutachter bzw. Erstgutachterin eingeteilt war. Andere möchten konzentriert in mehreren „Blöcken“ arbeiten und haben keine Zeit, um sich dazwischen die von anderen hinzugefügten Stellungnahmen

anzusehen. Wieder andere begutachten zwischen ihren anderen Tätigkeiten, was die Integration der PC-Tätigkeit einfach macht.

Die Interviews in der ersten Runde haben deutlich gemacht, dass es kein Problem ist, mit dem Begutachtungsprozess anzufangen. Das Kommentieren der Stellungnahmen von anderen Beteiligten ist schwieriger, da die PC-Mitglieder aufgrund der oben beschriebenen unterschiedlichen Arbeitsweisen unterschiedliche (Zeit-)Planungen haben. Das Kommentieren von anderen Beiträgen erfordert einen regelmäßigen Zugriff auf die Online-Plattform, was sich anscheinend schwierig in den Arbeitsalltag integrieren lässt. Wenn beispielsweise ein PC-Mitglied, Wochen nachdem es seine eigene Begutachtung des Antrags durchgeführt hat, auf Stellungnahmen oder Kommentare zu einem Antrag antworten soll, muss es möglicherweise den Antrag erneut lesen, was häufig zu zeitaufwendig ist.

Die Umfrage nach der zweiten Runde zeigt jedoch, dass die überwiegende Mehrheit es, wenn auch nur mäßig, leichter fand, die PC-Tätigkeit in bestehende (Zeit-)Planungen zu integrieren (Abb. 11). Es scheint einen Lerneffekt zu geben: In der zweiten Runde wird es etwas leichter. Die Umfrage zeigt, dass mehr als die Hälfte der Mitglieder zustimmt oder stark zustimmt, dass der PC mehr Flexibilität ermöglicht als der klassische Ansatz (Abb. 12).

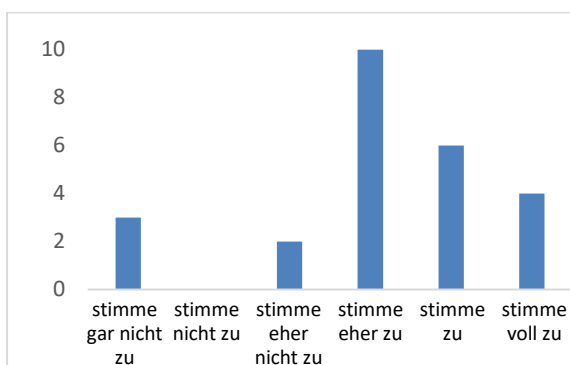


Abbildung 11: Die PC-Arbeit lässt sich einfach in meine normale Zeitplanung integrieren
N = 24

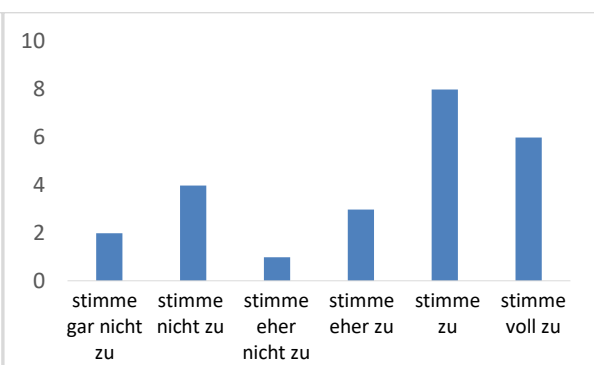


Abbildung 12: Der PC ermöglicht mehr Flexibilität als die klassische Begutachtung
N = 24

Ein wichtiger Punkt, der in den Interviews aufkam, ist, dass der Zeitraum für die Durchführung der Begutachtungen wahrscheinlich zu lang und noch nicht optimal strukturiert ist. Mehrere Befragte (Fachgutachter*innen, Ausschussmitglieder und AvH-Beschäftigte) schlugen vor, ihn zu verkürzen und/oder stärker zu strukturieren. Beispielsweise zwei Wochen für die anfänglichen Stellungnahmen und Kommentare und dann zwei Wochen zum wechselseitigen Kommentieren der ersten Beiträge. Ist den PC-Mitgliedern diese strukturierte (Zeit-)Planung im Voraus bekannt, können sie sie in ihre Terminplanung aufnehmen, außerdem würde es auch die Interaktion erleichtern. Und wenn es viele Anträge gibt, könnte ein solcher (Zeit-)Plan nochmals wiederholt werden.

Die Studie deckt zwei Runden des PC ab, da sich die Teilnehmer*innen an den PC gewöhnen und eine Routine entwickeln mussten. Beispielsweise wurde in der ersten Runde die Rolle des Erstgutachters bzw. der Erstgutachterin nicht von allen wirklich verstanden, in der zweiten Runde jedoch sehr viel besser. In den Interviews der Fachgutachter*innen und der Ausschussmitglieder haben wir sie gefragt, wie sie im Peer Circle vorgegangen sind, und es scheint sich ein „Durchschnittsverfahren“ abzuzeichnen (Textblock 1).

Schritt 1: In der Regel beginnt man mit den Anträgen, für die man als Erstgutachter bzw. Erstgutachterin eingeteilt ist. Diese Anträge wurden so begutachtet, wie es bei der klassischen Methode der Fall wäre, aber aufgrund des Formats des PC (Beantwortung von Fragen), musste keine „Geschichte“ geschrieben werden.

- Lesen des vollständigen Vorschlags und manchmal auch (von Teilen) der Veröffentlichungen. Die Veröffentlichungen wurden zumindest überflogen.
- Lesen und Beantworten der Kommentare der anderen PC-Mitglieder zu diesem Antrag.

Schritt 2: Die übrigen Anträge, die der Fachkompetenz des Fachgutachters bzw. der Fachgutachterin nahekommen.

- Zum Teil in Form einer vollständigen Begutachtung wie bei dem zugewiesenen Antrag, zum Teil eine weniger detaillierte Sichtung, aber Konzentration auf kritische Punkte.
- Nur wenige Peer Circle-Mitglieder (und einige Ausschussmitglieder) haben alle Anträge begutachtet/kommentiert.

Schritt 3: Schließlich die anderen Anträge, die weiter von der Fachkompetenz der Fachgutachter*innen entfernt sind. An dieser Stelle kommt es zur angestrebten Arbeitsteilung.

- Manche Fachgutachter*innen gehen alle Anträge durch, andere wählen einige aus, bei denen zusätzliche Kommentare sinnvoll scheinen, wobei sie häufig die Kommentare anderer PC-Mitglieder als Bezugspunkt verwenden. Zum Beispiel:
 - Vernachlässigung von Anträgen, für die bereits negative oder sehr positive Stellungnahmen vorliegen, und Schwerpunkt auf die Gruppe dazwischen; oder
 - Lesen und ggf. Beantwortung von Kommentaren anderer Teilnehmer*innen
- Diese Anträge werden kurz durchgesehen, der Schwerpunkt liegt dabei auf der Zusammenfassung und den für den AvH-Auswahlprozess wichtigen nicht fachlichen Teilen, wie Mobilität, Lebenslauf und Produktivität.
- Schließlich (Nicht-)Zustimmung zu einer Stellungnahme/einem Kommentar. Bei Nichtzustimmung Abgabe einer Begründung.

Textblock 1: Ein aufkommendes Verfahren

7.4 Zeitaufwand, Aufwand, Motivation

Einige der jüngeren PC-Mitglieder gaben in den Interviews an, nicht über Vorerfahrung mit Begutachtungen zu verfügen und konnten daher nicht beurteilen, ob der PC mehr oder weniger Zeit in Anspruch nimmt als der klassische Ansatz. Außerdem sind auf Selbstauskunft basierende Studien zum Zeitaufwand notorisch unzuverlässig. Da die meisten PC-Mitglieder online gearbeitet haben (Abb. 9), kann eine Analyse der Protokolldateien zusätzliche Erkenntnisse liefern.

Die Interviews zeigen, dass manche Fachgutachter*innen der Ansicht sind, der PC nehme mehr Zeit in Anspruch, vor allem weil aufgrund der Anzahl der Anträge der Leseaufwand höher ist. Andere denken, dass es weniger Zeit in Anspruch nehme, da wesentlich weniger Zeit zum Schreiben benötigt wird, was an der Darstellung der „Fragestruktur“¹⁴, dem informelleren Schreibstil und der Möglichkeit zur (Nicht-)Zustimmung zur Stellungnahme und zu Kommentaren anderer Teilnehmer*innen ohne Wiederholung der Argumente liegt.

Die Umfrage liefert ein differenziertes Bild. Ob das Lesen der Kommentare der anderen Mitglieder Zeit spart, ist nicht eindeutig; eine knappe Mehrheit stimmt zu, aber beinahe die Hälfte der PC-Mitglieder stimmt nicht zu (Abb. 13). Viele PC-Mitglieder empfinden das PC-

¹⁴ Das klassische HFST Peer Review-Verfahren nutzt auch die „Fragestruktur“, jedoch mit anderer Darstellung.

Verfahren pro Antrag als etwas zeitsparender (Abb. 14), Gleiches gilt für den PC als Ganzes (Abb. 15). Beinahe alle sind der Auffassung, dass sich der PC-Zeitaufwand innerhalb angemessener Grenzen bewegt (Abb.16).

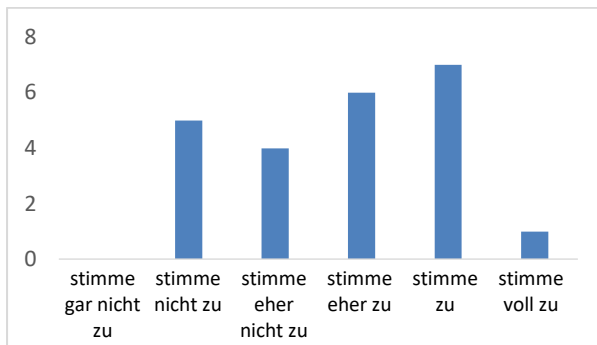


Abbildung 13: Das Lesen der Kommentare anderer Mitglieder spart Zeit
N = 23

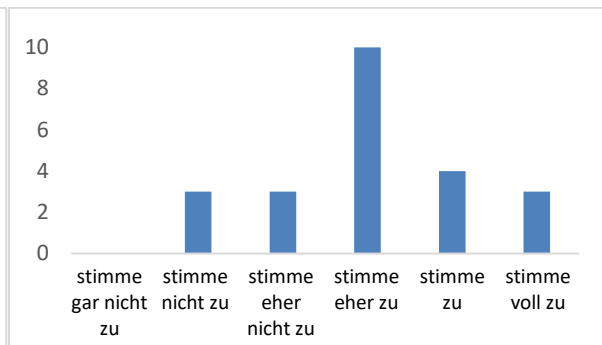


Abbildung 14: Der Zeitaufwand pro Antrag ist beim PC geringer
N = 23



Abbildung 15: Der Peer Circle spart Zeit
N= 22

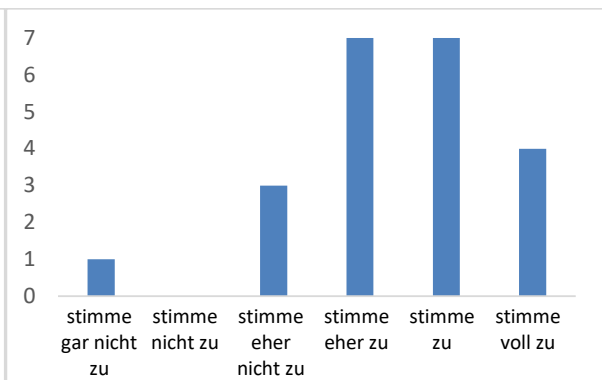


Abbildung 16: Der PC-Zeitaufwand ist angemessen
N = 22

Mehrere der Befragten, die zustimmen, dass der PC weniger Zeit pro Antrag benötigt, sind der Auffassung, dass die Teilnahme am PC insgesamt mehr Zeit in Anspruch nimmt. Im Gegensatz zur klassischen Peer Review-Methode, bei der die Fachgutachter*innen jeweils nur einen Antrag beurteilen, tragen die PC-Mitglieder zur Begutachtung mehrerer Anträge bei, was mehr Zeit in Anspruch nehmen würde. Die Befragten führten jedoch auch an, dass diese Methode für die *Forschungsgemeinschaft* als Ganzes effizienter sei. Es sind weniger Fachgutachter*innen als beim klassischen Ansatz beteiligt, und selbst wenn die PC-Mitglieder mehr Zeit aufwenden sollten, verringert sich der gesamte Zeitaufwand für die Begutachtung für die Gemeinschaft als Ganzes. In den vier Versuchsfachgebieten haben 28 PC-Mitglieder 89 Anträge begutachtet, wozu im klassischen Review-Verfahren 178 Fachgutachter*innen erforderlich gewesen wären.¹⁵

¹⁵ Basierend auf dem Grundsatz zwei unabhängiger Fachgutachter*innen pro Antrag. In einem Drittel bis zur Hälfte der Fälle wird diese Vorgabe nicht erfüllt, was die Probleme mit dem klassischen Peer Review-Prozess verdeutlicht. Wenn das Fachgebiet des Ausschussmitglieds dem Thema eines Antrags nahekommt, kann dieses auch als zweiter Gutachter bzw. zweite Gutachterin berufen werden, was den Bedarf an Fachgutachter*innen etwas reduziert. Gibt es nur einen Fachgutachter bzw. eine Fachgutachterin, entscheidet der Ausschuss, ob auf Basis dieses Fachgutachtens eine Entscheidung möglich ist.

Außerdem liefern die Protokolldateien Informationen zur für den PC aufgewandten Zeit. Da die meisten PC-Mitglieder hauptsächlich oder vollständig online gearbeitet haben, sind diese Dateien eine angemessene Quelle, auch wenn die Aktivitäten einiger Benutzer*innen nicht (vollständig) vom System erfasst wurden. Die Größe *Stunden eingeloggt* ist weniger zuverlässig, da hier auch die Zeit erfasst wird, in der ein Mitglied eingeloggt, aber mit etwas anderem beschäftigt ist. Die Kennzahl *Stunden aktiv* ist wesentlich aussagekräftiger, da sie die Zeit erfasst, in der die PC-Mitglieder an den Gutachten gearbeitet haben. Tabelle 11 zeigt die entsprechenden Informationen auf aggregierter Ebene. Die durchschnittliche aktive Online-Zeit beträgt rund 4 Stunden in der ersten (Sommer-)Runde und etwa halb so viel in der zweiten (Herbst-)Runde. Darin ist die Zeit für das Lesen nicht (immer) enthalten.

Tabelle 11: Login-Zeit auf der Online-Plattform

Alle Fachgutachter*innen und Ausschussmitglieder	Stunden eingeloggt		Stunden aktiv	
	Runde 1	Runde 2	Runde 1	Runde 2
Durchschnitt	25,9	14,7	4,2	2,1
Minimum	0,6	0,2	0,9	0,1
Maximum	77,1	49,7	9,4	7,9

Quelle: Protokolldateien

Einige Fachgutachter*innen haben in den Interviews versucht, zu schätzen, wie viele Tage sie für den PC aufgewandt haben. In der ersten Runde taten dies acht Fachgutachter*innen und die Zahl der Tage lag mit einem Durchschnitt von 4 Tagen bei 1 bis 8 Tagen. Gleiches gilt für einige Ausschussmitglieder, die den Zeitaufwand ebenfalls auf ein paar Tage geschätzt haben. Diese Befragten weisen in den Protokolldateien ein geringes „Nutzungs-niveau“ auf, was darauf hinweist, dass andere mehr Zeit für den Peer Circle aufgewandt haben.

Benutzerfreundlichkeit: Die Umfrage deutet darauf hin, dass es relativ einfach war, die Anträge auf der Online-Plattform zu kommentieren (Abb. 17). Der Durchschnittswert beträgt 4,5 und liegt damit zwischen „stimme eher zu“ und „stimme zu“ und etwas höher als nach der ersten Runde.

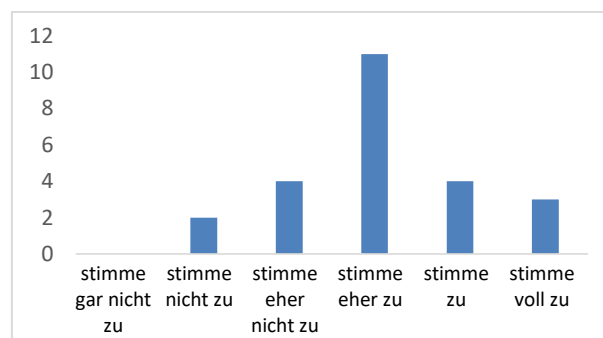


Abbildung 17: Es war einfach, die Vorschläge zu kommentieren
N = 24

Motivation: Viele Peer Circle-Mitglieder nahmen teil, weil sie (in der Vergangenheit) von der AvH-Stiftung gefördert wurden, ihre Teilnahme kann als Dienst für die Gemeinschaft und

genauer für die AvH-Stiftung angesehen werden. Dies ging aus den Interviews sowie aus der Umfrage hervor. Aber es gab auch andere Beweggründe:

- Mehrere Befragte wollten an dem Experiment teilnehmen, um einen Beitrag zur Verbesserung des Peer Review-Systems zu leisten.
- Andere, insbesondere weniger erfahrene PC-Mitglieder, erhofften sich zu lernen, wie die Begutachtung funktioniert und was aus Sicht der Fachgutachter*innen einen guten Antrag ausmacht. Sich mit vielen Anträgen zu befassen, war für sie sehr hilfreich.
- Außerdem nannten mehrere PC-Mitglieder Vorteile wie neue Ideen zu entwickeln, über Veränderungen im Fachgebiet auf dem Laufenden zu bleiben und die Kompetenzen im Bereich Antragsstellung zu verbessern.

7.5 Interaktion, Diskussion, Konsens

Interaktion & Diskussion: Die Interviews lassen annehmen, dass es nicht viel Interaktion gab, wenn auch in der Herbstrunde etwas mehr als in der Sommerrunde. Nach Ansicht der Befragten hat es durchaus einige Reaktionen auf den ersten Begutachtungsbeitrag auf der Plattform, aber kaum weiterführende Diskussionen gegeben. Die Umfrage zeigt, dass beinahe alle PC-Mitglieder einige oder regelmäßige Antworten festgestellt haben (Abb. 18).

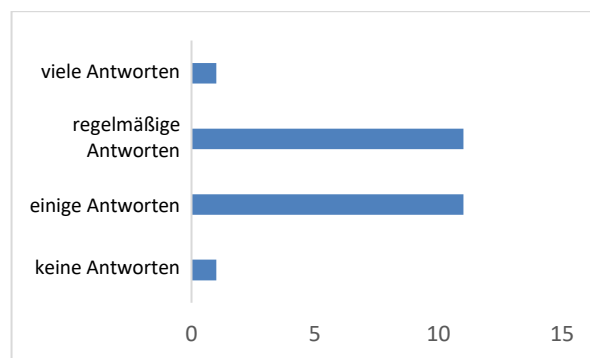


Abbildung 18. PC-Mitglieder haben auf die Kommentare anderer Mitglieder geantwortet
N = 24

Die Analyse der Textbeiträge auf der Online-Plattform kann weitere Erkenntnisse dazu liefern. Die Antworten auf die Begutachtungsfragen und die (anschließenden) Kommentare wurden mit einer fortlaufenden Nummer kodiert, die ihre Stelle in der Diskussion angibt und für die Erfassung der Anzahl der Diskussionen nach Länge verwendet wird. Eine inhaltliche Analyse der Begutachtungsbeiträge wäre sinnvoll, um die Daten zu bereinigen und unwesentliche Beiträge herauszunehmen, aber dies geht über den Rahmen dieses Berichts hinaus. Tabelle 12 zeigt die Ergebnisse. Wir haben hier alle Beiträge zum Austausch aufgenommen, auch die administrativen Mitteilungen der AvH-Beschäftigten, die etwa ein Drittel ausmachen. In Anbetracht dessen ist die Zahl der langen Diskussionen in der Tat bescheiden.

Tabelle 12: Länge der Diskussionen

Diskussionslänge*	Anzahl der Diskussionen	Summe der Beiträge
1	119	119
2	156	312
3	136	408
4	133	532
5	80	400
6	60	360
7	40	280
8	22	176
9	6	54
10	8	80
11	1	11
12	3	36

* Als Anzahl der Beiträge

Quelle: Protokolldateien der Online-Plattform

Konsens: Die PC-Mitglieder können dadurch beeinflusst werden, dass sie die Stellungnahmen und Kommentare der anderen Mitglieder lesen, ehe sie ihre eigene Begutachtung fertigstellen. In der ersten Runde lasen die meisten PC-Mitglieder überwiegend zuerst die Kommentare der anderen Mitglieder. Dies war in der zweiten Runde anders: Die meisten Fachgutachter*innen gaben an, in der zweiten Runde nicht systematisch angefangen zu haben, die Stellungnahmen und Kommentare der anderen Mitglieder zu lesen, ehe sie ihre eigenen erstellt haben (Abb. 19). Dies lässt annehmen, dass die meisten PC-Mitglieder eine eigene Einschätzung haben, die sie durch Kommentare zu den Stellungnahmen anderer Mitglieder teilen (Abb. 20) – was eine der Ideen hinter dem PC-Konzept ist.

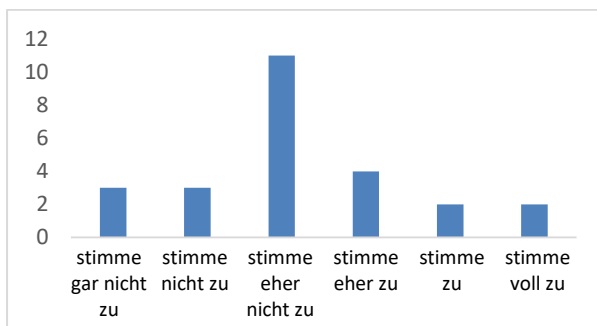


Abbildung 19: Ich habe die anderen Kommentare gelesen, ehe ich mit meinen eigenen begonnen habe
N = 25

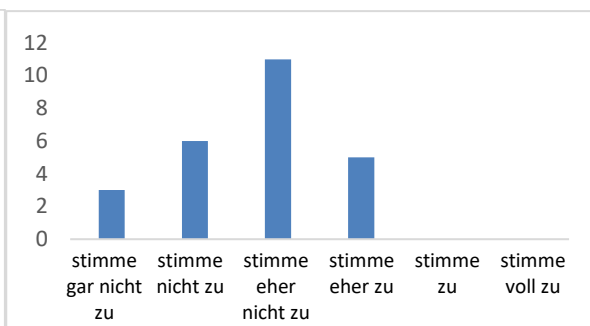


Abbildung 20: Ich habe die Beurteilungen anderer Mitglieder kommentiert, statt eine eigene zu erstellen
N = 25

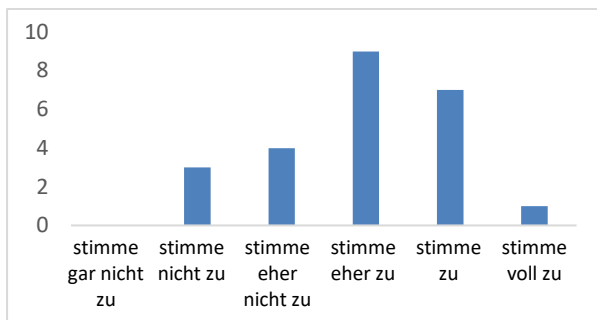


Abbildung 21: Ich wurde durch die Kommentare anderer Mitglieder beeinflusst
N = 24

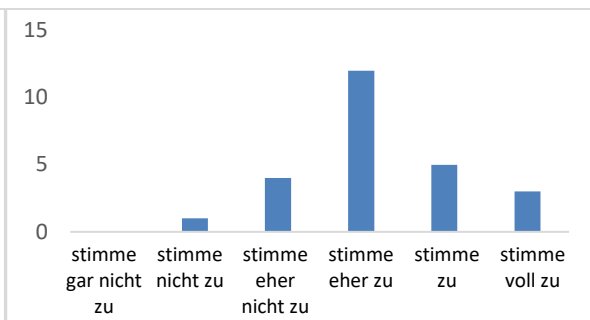


Abbildung 22: Das Lesen der Kommentare anderer Mitglieder half mir bei meiner Meinungsbildung
N = 25

Die meisten Befragten gaben an, zu einem gewissen Grad durch das Lesen der Kommentare anderer Fachgutachter*innen beeinflusst worden zu sein (Abb. 21). Die Interviews ergaben das gleiche Bild: Beinahe allen befragten Fachgutachter*innen war die Möglichkeit einer Beeinflussung in einem Gruppenprozess wie dem PC bewusst und mehrere von ihnen gaben an, sie seien geneigt gewesen, sich den Kommentaren der anderen Mitglieder, insbesondere der jeweiligen Erstgutachter*innen, anzuschließen. Diese Befragten äußerten jedoch auch, sie würden dies nicht aus den Augen verlieren und reflektieren, ob dies der Fall sei.

Sie sind auch der Ansicht, dass die Beeinflussung durch andere Vorteile habe, weil sie helfe, die eigene Meinung zu bilden (Abb. 22) und zu reflektieren, was letztlich zu einer anderen Einschätzung führen könne. Ein weiterer Vorteil der wechselseitigen Beeinflussung besteht darin, dass sie bei der Konsensfindung hilft. Dies ist positiv, da ein Ziel des Begutachtungsprozesses darin besteht, zu einem Konsens über die zu fördernden Antragsteller*innen zu gelangen.

Beeinflusst zu werden ist nicht das Gleiche wie ein *vorzeitiger Konsens* mit der Gefahr einer Unterdrückung etwaiger abweichender Ansichten. Den Befragten (Fachgutachter*innen, Ausschussmitglieder und Fächergruppenleitungen) war dieses Risiko bewusst, aber so gut wie niemand von ihnen war der Ansicht, dass es eingetreten war. Es wurden ein oder zwei Fälle erwähnt, in denen sich ein früher Konsens zu bilden schien, aber diese wurden im PC ausdrücklich angesprochen.

7.6 *Interaktion zwischen Ausschussmitgliedern und Fachgutachter*innen*

Im Peer Circle Neuere Geschichte gab es keine Interaktion zwischen Fachgutachter*innen und Ausschussmitgliedern, da die Ausschussmitglieder erst am Ende der Peer Circle-Tätigkeiten in den Prozess einstiegen. Auch eines der Ausschussmitglieder im Fachgebiet Materialwissenschaft stieg erst nach Abschluss der PC-Tätigkeiten ein. Aus Sicht der Fachgutachter*innen stellte die Anwesenheit von Ausschussmitgliedern im Peer Circle kein Problem dar. Auch wenn einige Ausschussmitglieder stark als Gutachter*innen aktiv waren, war das den übrigen Gutachter*innen möglicherweise nicht einmal bewusst, da alle PC-Mitglieder anonym waren.

Die Ausschussmitglieder verwenden die Ergebnisse der Fachgutachter*innen und integrieren sie in einen Vorschlag für den Ausschuss. Mögliche Spannungen können dadurch auftreten, dass sich im PC nicht nur das Ausschussmitglied, sondern auch einige oder alle PC-Mitglieder einen guten Überblick über den Satz von Anträgen verschafft haben. Bisher ist es Aufgabe des Ausschussmitglieds, einen Vorschlag für den Ausschuss zu erstellen, aber dies könnte zunehmend auch der Peer Circle übernehmen, da der Informationsvorsprung des Ausschussmitglieds abnimmt. Mehrere PC-Mitglieder deuteten dies an, indem sie vorschlugen, der PC könne seine eigene Rangfolge erstellen. Ein Ausschussmitglied schlug vor, zunächst das Feedback des PC zur vorgeschlagenen Rangfolge einzuholen, ehe sie dem Ausschuss vorgelegt wird.

7.7 Vergleich mehrerer Anträge

Ein Nachteil des klassischen Peer Review-Verfahrens besteht darin, dass die Fachgutachter*innen jeweils nur ein einziges Fachgutachten erstellen und daher nicht die Qualität des Antrags im Vergleich zu den anderen Anträgen beurteilen können. Dies führt häufig dazu, dass alle Anträge, unabhängig von der Menge an Lob und Kritik in der gutachterlichen Stellungnahme, relativ gut benotet werden. Im PC sehen die Mitglieder alle Anträge und können die verschiedenen Anträge daher untereinander vergleichend benoten. Beinahe alle PC-Mitglieder sehen dies als Vorteil an. Die Interviews weisen in dieselbe Richtung, und viele der Befragten sind der Ansicht, dass ein Vergleich der Anträge die Begutachtung einfacher (Abb. 23), aber auch besser macht. Das Ergebnis ist eine bessere Vorbereitung für die Ausschussmitglieder, die sich so auf eine wesentlich breitere vergleichende Benotung der Anträge stützen können.

Wäre es hilfreich, wenn ein PC eine gemeinsame Rangfolge der Anträge erstellen würde? Die Meinungen gehen auseinander, etwas mehr als die Hälfte der Befragten stimmt tendenziell zu, die etwas kleinere andere Hälfte stimmt tendenziell nicht zu (Abb. 24). Mehrere Befragte sprachen den Vergleich und die Bildung einer Rangfolge der Anträge im PC und die praktischen Schwierigkeiten bei der Organisation der Rangfolgenerstellung an. In diesem Zusammenhang wurde auch darauf hingewiesen, dass die Fachgutachter*innen die komplette Vier-Punkte-Skala und nicht nur 2, 3 und 4 nutzen sollten, weil dies eine bessere Differenzierung der Anträge ermöglichen würde.

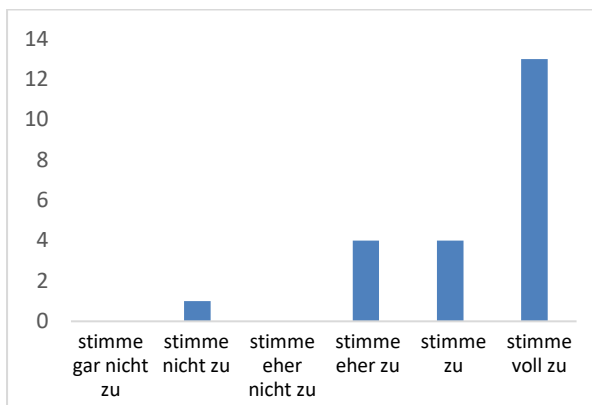


Abbildung 23: Der Vergleich der Anträge macht die Begutachtung einfacher
N = 22

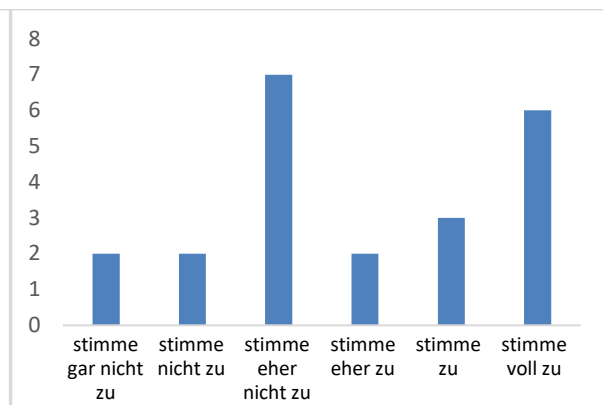


Abbildung 24: Der Peer Circle sollte eine eigene Rangfolge entwickeln
N = 22

7.8 Akzeptanz

Das Experiment zeigt, dass der PC von der Wissenschaftsgemeinde gut aufgenommen werden könnte. In den Interviews äußerte sich eine Mehrheit der PC-Mitglieder positiv über den Peer Circle als Alternative zum klassischen Peer Review-Modell und gab an, er sei die Lösung für die Zukunft. Die Umfrage (Abb. 25) zeigt, dass 18 der 24 Befragten eine positive bzw. sehr positive und eine weitere Person eine etwas positive Ansicht vertraten. Nur drei vertraten eine etwas negative Ansicht. Wenn wir zwischen den vier Fachgebieten unterscheiden, sind gewisse, aber keine großen Unterschiede festzustellen. Drei PC weisen recht gute Wertungen auf: Zoologie, Neuere Geschichte und Anorganische Chemie liegen mit

Durchschnittswerten von 5,2, 5,1 bzw. 5,0 nahe beieinander. In der Materialwissenschaft ist die Wertung mit 4,4 auf einer Sechs-Punkte-Skala etwas niedriger, aber immer noch positiv.

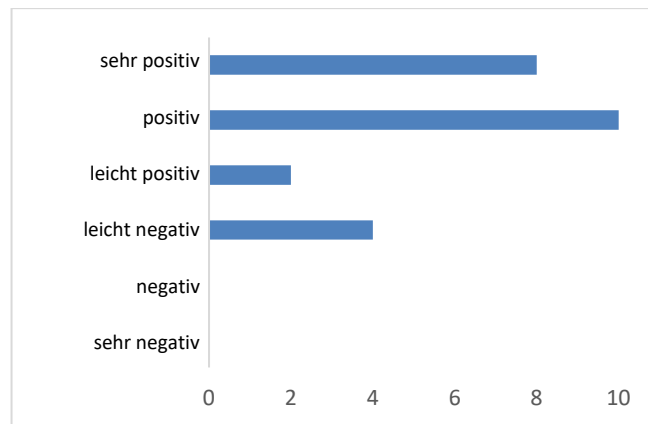


Abbildung 25: Allgemeiner Eindruck des PC-Verfahrens
N = 24

Die Interviews ergaben einen Bedarf an neuen Lösungen für Probleme im Peer Review-System und die meisten Ausschussmitglieder bevorzugten für das HFST-Programm das PC-Konzept. Einige führten auch Nachteile an, insbesondere, dass es für ein Ausschussmitglied im PC schwieriger sei, vom Konsens des PC abzuweichen. Die einzige negative Einschätzung kam von einem Ausschussmitglied im Fachgebiet Neuere Geschichte, welches der Ansicht ist, dass Stellungnahmen von Fachgutachter*innen ohne passende Spezialisierung im Auswahlprozess nicht hilfreich seien.

8. Kontext und Implementierung des Verfahrens

8.1 Peer Circle-Mitglieder

Eines der Ziele des PC ist eine vielfältigere Zusammensetzung, so sollen ihm u. a. auch Fachgutachter*innen in früheren Karrierestufen angehören, die noch keine Professur haben. Tabelle 13 zeigt die Zusammensetzung der Peer Circle und es gibt einige Unterschiede zwischen den vier Fachgebieten. Das Ziel, mehr Frauen und jüngere Wissenschaftler*innen in den PC einzubeziehen, wurde aber erreicht. Beinahe die Hälfte der PC-Mitglieder sind Frauen und die Altersverteilung ist mit mehr als einem Drittel der PC-Mitglieder im Alter von unter 40 gleichmäßig.

Tabelle 13: Demografie des Peer Circle

	Frauen	Alter 30-39	Alter 40-49	Alter 50-59	über 60	Prof.	Summe
Peer Circle-Mitglieder	43 %	36 %	32 %	31 %		46 %	28
Ausschussmitglieder	14 %			71 %	29 %	100 %	7

Wie viele Anträge haben die 28 PC-Mitglieder begutachtet? In der ersten Runde waren es 50 und in der zweiten Runde 42. Zoologie hatte mit 30 Anträgen den größten Anteil, es folgten Neuere Geschichte, Anorganische Chemie und Materialwissenschaft mit 20, 20 bzw. 19 Anträgen (Tabelle 14).

Tabelle 14: Anzahl der Anträge pro Runde

Peer Circle	Runde 1	Runde 2
Anorganische Chemie	10	10
Materialwissenschaft	10	9
Zoologie	16	14
Neuere Geschichte	12	8
Summe	48	41

8.2 Anonymität

In der Vorbereitungsphase wurde diskutiert, ob die Peer Circle-Mitglieder anonym sein sollten. Die Idee dahinter war, dass sich insbesondere die weniger erfahrenen Mitglieder ohne Festanstellung weniger frei fühlen würden, ihre Meinung zu äußern, wenn diese der Bewertung der erfahreneren Mitglieder des PC widerspricht. Vor Beginn der ersten Runde stimmten die PC- und Ausschussmitglieder mehrheitlich dafür, den Prozess anonym zu halten. Es herrschte keine vollständige Anonymität, da sich einige Teilnehmer*innen bei einer der zwei Einführungsveranstaltungen, bei denen das Verfahren und die Online-Plattform vorgestellt wurden, (online) kennengelernt hatten. Nach Abschluss des PC sollten die Ausschussmitglieder die Identität der Fachgutachter*innen erfahren. Dies wurde zwar bei der Einführungsveranstaltung und in dem von der AvH bereitgestellten Informationsmaterial erwähnt, schien aber nicht allen Fachgutachter*innen bewusst zu sein.

Die Frage der Anonymität wurde in den Interviews angesprochen und die Hauptargumente für Anonymität sind, (i) dass sie jüngere, insbesondere befristet beschäftigte Wissenschaftler*innen schützt, sodass sie ihre Meinung frei äußern können, (ii) dass Begutachtungen nicht

durch Statusunterschiede behindert werden sollten, (iii) dass sie die Vertraulichkeit schützt und (iv) dass sie die Wahrscheinlichkeit einer Beeinflussung von Fachgutachter*innen durch andere Mitglieder verringert. Einige Befürworter*innen von Anonymität argumentierten, dass während des PC-Prozesses alle anonym bleiben, anschließend die Namen der Fachgutachter*innen jedoch offengelegt werden sollten (was für die Ausschussmitglieder bereits der Fall ist). Dies würde es einfacher machen, die Kommentare zu gewichten und zu verhindern, dass *Befangenheiten* aufgrund von Anonymität unbemerkt bleiben. Laut einem*r AvH-Beschäftigten gibt es allerdings bereits an mehreren Punkten im Verfahren Überprüfungen auf Befangenheiten – was eventuell nicht allen PC-Mitgliedern bekannt ist.

Die Gegner*innen von Anonymität führen an, dass (i) auch junge Wissenschaftler*innen in der Lage sein sollten, ihre Ansichten öffentlich zu verteidigen (wie auf Konferenzen), (ii) Anonymität zu möglicherweise unbeobachteten *Befangenheiten* führt, (iii) Anonymität die Interaktion erschwert, (iv) Anonymität die Gewichtung von Kommentaren unmöglich macht und (v) Anonymität die Transparenz verringert.

Aus den Interviews geht hervor, dass die Mehrheit der PC-Mitglieder Anonymität befürwortet, und die Umfragen zeigen, dass die Mehrheit zwischen den beiden Runden etwas gewachsen ist: von zwei Drittel in der ersten Befragung auf drei Viertel in der zweiten Befragung. Diejenigen, die sich gegen Anonymität aussprachen, waren zumeist nur leicht dagegen, d. h. „etwas gegen Anonymität“ (Abb. 26). In der Gruppe der Ausschussmitglieder war eine stärkere Veränderung festzustellen. Nach der ersten Runde waren drei von sieben Ausschussmitgliedern gegen Anonymität, nach dem zweiten Interview nur noch eines von acht.

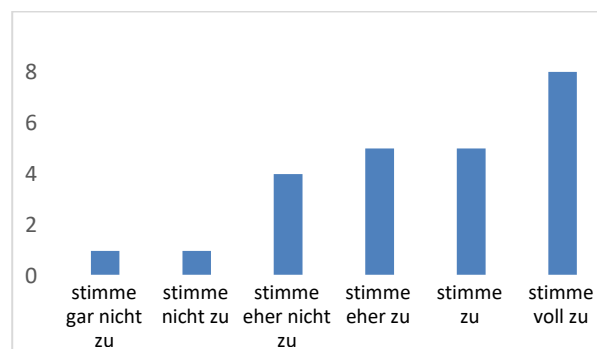


Abbildung 26: Anonymität ist für den PC von Vorteil
N = 24

8.3 Allgemeine Aspekte des Verfahrens

*Erstgutachter*in:* Um zu verhindern, dass alle warten, bis jemand anderes mit der Begutachtung anfängt, haben die Koordinator*innen für jeden Antrag ein oder zwei PC-Mitglieder, deren Fachgebiet dem des Antrags nahekommt, aufgefordert, mit dem Begutachtungsprozess zu beginnen. Die Interviews in der ersten Runde zeigen, dass nicht alle PC-Mitglieder die Rolle des Erstgutachters bzw. der Erstgutachterin eindeutig verstanden hatten, insbesondere nicht, dass von ihnen erwartet wurde, früh mit dem Begutachtungsprozess zu beginnen, was in manchen PC zu einem späten Start führte. In der zweiten Runde wurden die Rollen besser verstanden (Abb. 27) und es war einfach, früh anzufangen (Abb. 28). Dennoch zeigen die Protokolldateien, dass in zwei PC ein später Beginn weiterhin ein Problem darstellte.

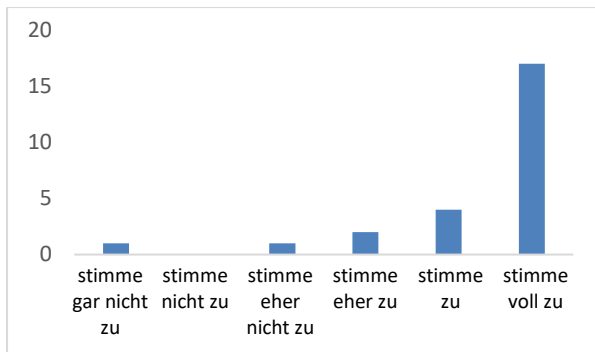


Abbildung 27: Wenn ich als Erstgutachter*in eingesetzt war, fühlte ich mich angehalten, die Begutachtung zu beginnen
N = 25

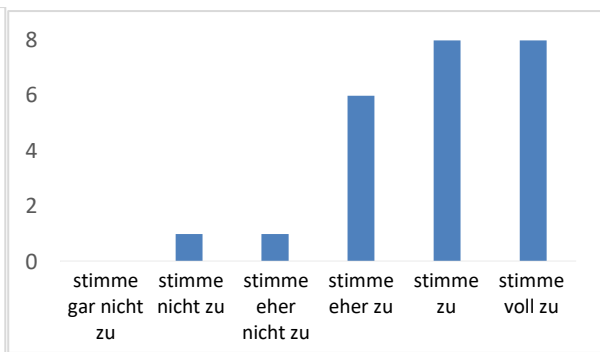


Abbildung 28: Ich fand es einfach, als Erste(r) mit einer Begutachtung zu beginnen
N = 25

*Gastgutachter*innen:* Gelegentlich war ein Projektvorschlag nicht von der Fachkompetenz der PC-Mitglieder abgedeckt und es wurde ein Gastgutachter bzw. eine Gastgutachterin eingeladen. Laut Angaben der Fächergruppenleitungen waren diese nicht schwer zu finden und die Fachgutachter*innen hielten dies für eine praktikable Lösung. Die Ausschussmitglieder wiesen auch darauf hin, dass es in diesen Fällen erforderlich war, externe Fachgutachter*innen einzuladen. Die fehlende Übereinstimmung von Fachgebieten ist jedoch kein spezifisches Problem des PC und mehrere Befragte führten an, dass es auch im klassischen Peer Review-Prozess auftrete. Im Peer Circle ist es einfacher zu erkennen als bei klassischen Begutachtungen.

Phasen im PC-Prozess: Der PC dauerte recht lange: die erste Runde elf Wochen und die zweite Runde acht Wochen. Um die Interaktion zwischen den Fachgutachter*innen zu fördern, wurden in jeder Runde zwei zweiwöchige „interaktive Zeiträume“ angesetzt. Die meisten PC-Mitglieder fanden die interaktiven Zeiträume mäßig hilfreich (Abb. 29) und angesichts der Werte für die Frage „Ich hatte den Eindruck, dass ich in den interaktiven Zeiträumen begutachten sollte“ war die Rolle der interaktiven Zeiträume nicht allen Fachgutachter*innen klar: Nur die Hälfte der PC-Mitglieder stimmte dem zu (Abb. 30). Die Analyse der Protokolldateien (Übersicht 1 und 2, Abschnitt 7.1) weist darauf hin, dass sich die Begutachtungstätigkeiten in der Regel auf den Anfang und das Ende der Peer Circle-Zeiträume konzentrierten und nicht immer in den interaktiven Zeiträumen lagen. Beispiele hierfür sind die Zoologie mit einem abweichenden Muster in beiden Runden sowie die Anorganische Chemie und die Materialwissenschaft in der zweiten Runde.

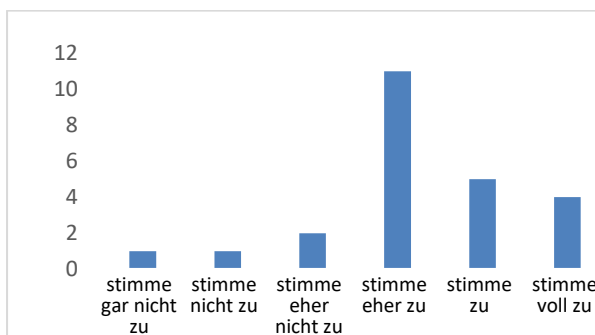


Abbildung 29: Die interaktiven Zeiträume waren hilfreich
N = 24

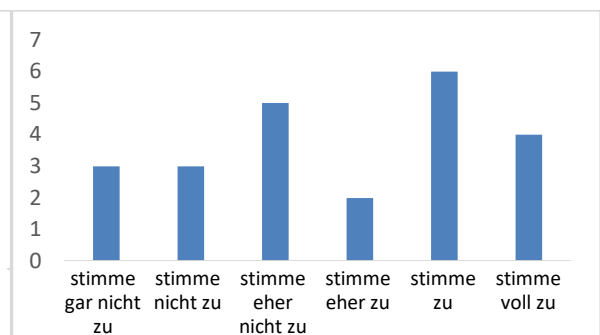


Abbildung 30: Ich hatte den Eindruck, dass ich nur im interaktiven Zeitraum begutachten sollte
N = 23

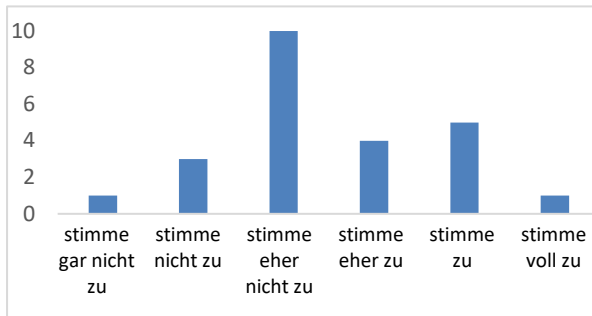


Abbildung 31: Ich habe vor, meine Aktivitäten über längere Zeit zu verteilen
N = 24

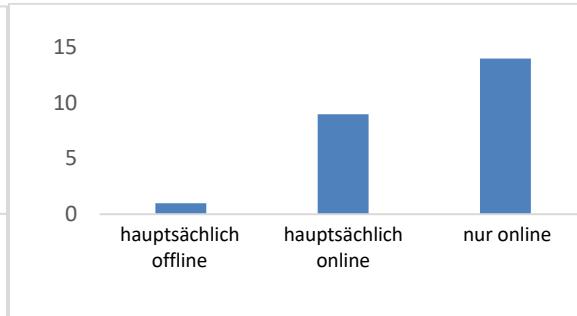


Abbildung 32: Wie haben Sie die Stellungnahmen/Kommentare erstellt
N = 24

Die Interviews haben diese Ergebnisse mit den unterschiedlichen Arbeitsweisen der PC-Mitglieder in Zusammenhang gebracht. Nicht wenige der Befragten äußerten, dass die PC von kürzerer Dauer und stärker strukturiert sein sollten, mit unterschiedlichen Fristen für die ersten Stellungnahmen und die Kommentare.

Benotung: Die PC-Mitglieder sollten nur eine Gesamtnote für die Anträge und keine Einzelnote für jedes Bewertungskriterium vergeben. Die Befragten sprachen dieses Thema an, das als wichtig, aber z. T. auch als schwierig angesehen wird: Wie soll ein Antrag insgesamt benotet werden, wenn sich die einzelnen Fachgutachter*innen nicht mit allen Aspekten des Antrags befasst haben? Die Benotung einzelner Bewertungskriterien (wie beim klassischen Ansatz) wäre gerade für den Peer Circle besser geeignet.

Auch die 4-Punkte-Skala wurde angesprochen. Da in der Praxis nur ein Teil der Skala genutzt wird, hielten mehrere Befragte sie für nicht nuanciert genug für einen angemessenen Vergleich der Anträge. Dies ist ein wichtiger Punkt, insbesondere da die vergleichende Beurteilung einer der Hauptvorteile des PC ist.

Online- gegenüber Offline-Tätigkeit: Beinahe alle PC-Mitglieder haben vollständig oder hauptsächlich online gearbeitet (Abb. 32), sodass die Protokolldateien ein sinnvoller Datensatz zur Analyse der Aktivitäten der PC-Mitglieder sind (s. Abschnitt 7.2). Einige Fachgutachter*innen regten an, Offline-Arbeit mit automatischem Upload auf die Plattform zu ermöglichen. Hintergrund dieser Anregung ist, dass wenn jemand Zeit für die Begutachtung hat, nicht immer ein Internetzugang verfügbar ist (z. B. unterwegs).

Einladung der PC-Mitglieder: Die Suche nach Mitgliedern für den PC schien laut den Interviews mit den AvH-Beschäftigten und den Ausschussmitgliedern einfacher als die Suche nach klassischen Fachgutachter*innen. Dies lag zum Teil daran, dass die eingeladenen PC-Mitglieder alle aus dem AvH-Netzwerk stammten, aber vor allem ist die Zahl der PC-Mitglieder deutlich geringer als die Zahl der klassischen Fachgutachter*innen, die für die gleiche Menge an Antragsteller*innen benötigt würden.

Diesbezüglich führten einige Befragte an, dass die Fachgutachter*innen zwei oder drei Jahre lang PC-Mitglieder bleiben sollten, um ausreichend Erfahrung zu sammeln. Dies würde auch Zeit bei der Suche nach Fachgutachter*innen sparen. Um starke Schwankungen bei der im PC verfügbaren Erfahrung zu verhindern, sollten nicht alle PC-Mitglieder gleichzeitig ausgetauscht werden.

Schließlich ist die Sprache der Plattform, ebenso wie die der meisten Stellungnahmen und Kommentare, Deutsch. Dies beeinträchtigt die Teilnahme von nicht-deutschsprachigen Personen.

Vorbereitung der Materialien: Der PC ist ein Experiment und daher (noch) nicht in die digitalen Standardprozesse der AvH eingebunden. Dies bedeutet, dass die Vorbereitung der Dokumente für den PC und die Online-Plattform in einem anderen als dem üblichen Format für die Fachgutachter*innen mit einem erheblichen Arbeitsaufwand einhergeht. Aus den Interviews mit den Sachbearbeitungen wird deutlich, dass in der aktuellen Situation für den PC ein beträchtliches administrative Arbeitspensum anfällt. Sie gehen jedoch davon aus, dass dieses Problem durch die Integration einer dedizierten PC-Plattform in den normalen Arbeitsablauf beseitigt wird. Bis dahin erfordert der PC einen Verwaltungsmehraufwand.

Koordinierung des Peer Circle als neue Aufgabe: Beim klassischen Begutachtungsverfahren ist, abgesehen davon, die Fachgutachter*innen zur rechtzeitigen Abgabe der Gutachten anzuhalten und nach Einreichung des Gutachtens die Qualität zu überprüfen, keine Intervention und Koordination erforderlich. Im klassischen Begutachtungsverfahren gibt es keine Interaktion mit den Fachgutachter*innen, aber im PC können die AvH-Beschäftigten und die Ausschussmitglieder Erläuterungen zu unklaren Punkten einholen und die PC-Mitglieder auffordern, sich bestimmte Dinge noch einmal anzusehen. Allgemeiner gesagt, ist der PC eine Gruppentätigkeit mit höherer Interdependenz und größerem Koordinierungsbedarf. Im Rahmen ihres Koordinierungsauftrags mussten sich die AvH-Beschäftigten recht häufig auf der Plattform einloggen, die Aktivität oder deren Ausbleiben überwachen und die PC-Mitglieder auffordern, mit der Begutachtung zu beginnen oder noch nicht abgedeckte spezifische Themen zu behandeln. Auch die Beantwortung von Fragen der PC-Mitglieder gehörte zur Koordination. Dies führt zu einer zusätzlichen Arbeitsbelastung, die möglicherweise zurückgeht, wenn sich die Fachgutachter*innen an die neue Arbeitsweise gewöhnen, aber es ist und bleibt eine neue Aufgabe für die AvH-Beschäftigten. Anders gesagt, der PC bedeutet veränderte Rollen und Aufgaben der beteiligten AvH-Beschäftigten und eine Veränderung des Arbeitsablaufs.

Die Interviews (mit Fachgutachter*innen und Ausschussmitgliedern) zeigen, dass die Koordinierungsaufgaben der AvH-Beschäftigten hochgeschätzt und als für das Funktionieren des Peer Circle wesentlich angesehen wurden. Dies zeigt sich auch in der Umfrage: 21 von 24 PC-Mitgliedern stimmten (stark) zu, dass die Moderation durch die AvH-Geschäftsstelle hilfreich war (Abb. 33) und in Bezug auf die Kommunikation mit der AvH-Geschäftsstelle vertraten 20 von 25 PC-Mitgliedern diese Ansicht (Abb. 34).

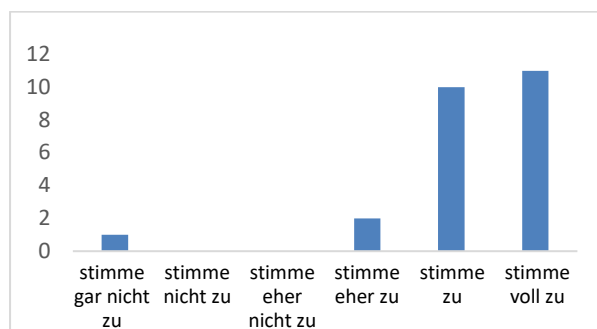


Abbildung 33: Die Moderation der AvH war hilfreich
N = 24

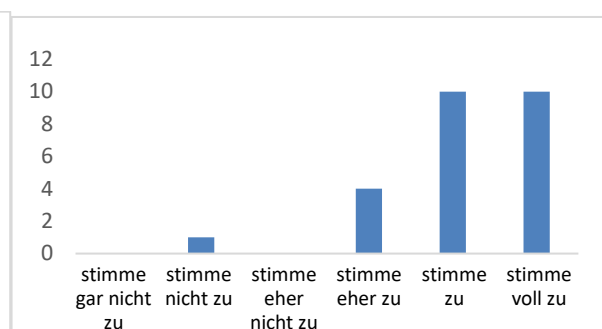


Abbildung 34: Die Kommunikation mit der AvH war hilfreich
N = 25

Zusammenfassender Bericht: Die Interviews mit den Fächergruppenleitungen weisen darauf hin, dass die Erstellung des zusammenfassenden Berichts einfacher ist, da (i) die Fächergruppenleitungen aktiver am Begutachtungsprozess beteiligt sind und (ii) bei Unklarheiten einfacher Rückfragen an die Fachgutachter*innen möglich waren.

8.4 Funktionalität und Benutzerfreundlichkeit

Der Peer Circle wurde auf einer bestehenden Online-Plattform, die nicht in die technischen Verfahren der AvH integriert ist, implementiert; wenn der PC fortgesetzt wird, wird eine dedizierte Plattform benötigt. Die aktuelle Evaluation zielt daher nicht auf die Bewertung der Benutzerfreundlichkeit der Online-Plattform ab. Wir haben jedoch in der ersten Umfrage und in den Interviews danach gefragt, um herauszufinden, ob sich Probleme bei der Benutzerfreundlichkeit auf das Experiment ausgewirkt haben. Auch wenn manche Benutzer*innen zum Teil Probleme hatten, schien dies nicht der Fall zu sein. Die meisten PC-Mitglieder¹⁶ hielten die Plattform für in Ordnung und benutzerfreundlich: intuitiv, einfach zu nutzen und selbsterklärend. Die Interviews weisen in dieselbe Richtung, und von den zwölf Befragten in der ersten Runde hatten neun eine positive Ansicht zur Online-Plattform und drei waren eher gemischter Ansicht. Dies hat sich nach der zweiten Runde nicht geändert. Die Ausschussmitglieder waren, im ersten Interview stärker als im zweiten, kritischer als die Fachgutachter*innen. Viele Befragte führten Verbesserungsvorschläge für die Gestaltung einer dedizierten Peer Circle-Plattform an. Zwei dieser Punkte sollen hier angesprochen werden:

Zum einen fanden es (in der ersten Runde) nur etwa die Hälfte bis zwei Drittel der Fachgutachter*innen relativ leicht, sich einen Überblick über den Status individueller Begutachtungen zu verschaffen und zu erkennen, wann ein Antrag ausreichend begutachtet worden war. Diese Meinungen können sich mit zunehmender Erfahrung der Fachgutachter*innen ändern, aber bei der Herbstumfrage konnte nur ein Unterschied für „Leichtigkeit, den Überblick zu behalten“ festgestellt werden. Wesentlich mehr Befragte stimmen zu einem gewissen Grad zu, dass dies leicht ist (Abb. 35). Jedoch ist es für recht viele PC-Mitglieder weiterhin unklar, wann ein Antrag angemessen begutachtet wurde (Abb. 36).

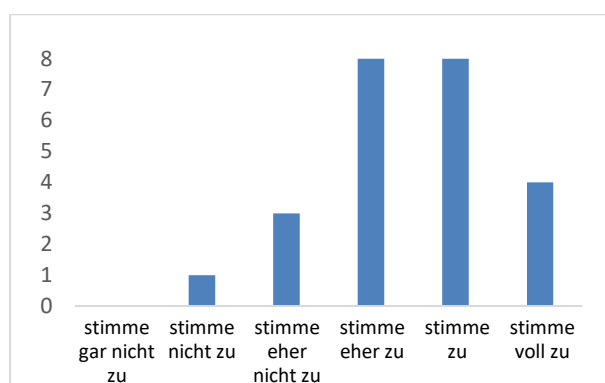


Abbildung 35: Es war leicht, den Überblick über den Begutachtungsprozess zu behalten
N = 24

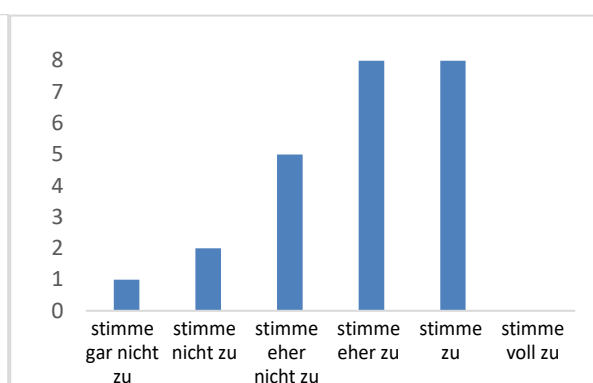


Abbildung 36: Es war immer klar, wann ein Antrag ausreichend begutachtet war
N = 24

¹⁶ Elf von dreizehn in der ersten Umfrage.

Ein weiterer Punkt ist die Art und Weise, wie die Dokumente auf der Online-Plattform organisiert sind. In der ersten Runde stimmte ein Drittel eher nicht zu, dass die Unterlagen im PC besser organisiert waren als bei klassischen Begutachtungsformen. Nach der zweiten Runde hatte sich dies etwas verändert: Rund 80 % gaben eine positive Bewertung ab (Abb. 37). Der Durchschnittswert blieb jedoch gleich (3,9 in beiden Umfragen), was anzeigt, dass die kritischen PC-Mitglieder in der zweiten Umfrage sogar noch niedrigere Wertungen abgaben. Zu diesem Punkt fehlen einige Werte, aber die Teilnehmer*innen, die nicht geantwortet haben, unterschieden sich nicht nach Alter oder Forschungsgebiet von denjenigen, die geantwortet haben.

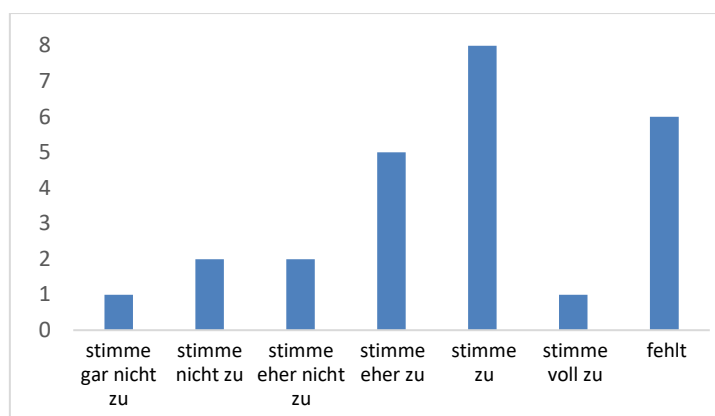


Abbildung 37: Die Unterlagen waren besser organisiert
N = 19

Teil III – Schlussfolgerungen und Empfehlungen

9. Schlussfolgerung

Mit dieser Evaluation sollen die folgenden Fragen zum Peer Circle beantwortet werden:

1. Sind die PC-Gutachten den klassischen Fachgutachten qualitativ mindestens ebenbürtig?
2. Führt die Interaktion im PC zu einem vorzeitigen Konsens, der eine kritische Betrachtung der Anträge verhindert?
3. Sehen die Fachgutachter*innen und Ausschussmitglieder den PC als bessere Alternative an?
4. Ist der PC, etwa in Bezug auf den Zeitaufwand, effizienter als der klassische Ansatz?
5. Wählt der PC-Ausschuss die besten Antragsteller*innen aus?
6. Wirken sich PC-Stellungnahmen auf das Geschlechtergleichgewicht im Auswahlresultat aus?

1. Qualität der Gutachten

Das Format der PC-Gutachten unterscheidet sich bewusst von dem klassischer Gutachten: Es ist kürzer, informeller und enthält weniger eine konstruierte Story, die zu einem Urteil führt. Dies ergibt sich dadurch, dass nur Stellungnahmen zu den Teilen des Antrags eingeholt werden, mit denen die jeweiligen Fachgutachter*innen vertraut sind, sowie durch das interaktive Online-Format, in dem die Fachgutachter*innen ihre Beiträge untereinander kommentieren können. Wir haben auch festgestellt, dass der Schwerpunkt weniger auf den fachlichen Detailfragen des vorgeschlagenen Projekts liegt, was laut den meisten Befragten ein Vorteil ist, da der Schwerpunkt bei der Begutachtung auf der Gesamtleistung der Person und nicht auf dem Projekt liegen sollte. Dies verhindert eine zu starke Konzentration auf fachliche Einzelheiten, die von der Gesamtqualität des Antrags ablenkt. Nur wenige der Befragten hielten dies für einen Nachteil des PC.¹⁷

Die meisten Teilnehmer*innen des Experiments halten die Summe der Stellungnahmen und Kommentare für den klassischen Gutachten qualitativ mindestens ebenbürtig. Gleiches gilt für die Vollständigkeit: Bei der PC-Begutachtung wurden alle Aspekte der Anträge abgedeckt.

Besteht die Gefahr, dass bei manchen Anträgen keines der PC-Mitglieder über die erforderliche Fachkompetenz verfügt? Laut den meisten Befragten wird dieses mögliche Problem durch die Rolle der „Erstgutachter*innen“ und die Möglichkeit zur Einbeziehung weiterer Gutachter*innen in den Peer Circle für bestimmte (Teile von) Anträgen gelöst.

Als Hauptvorteil wird genannt, dass im Peer Circle eine größere Gruppe von Fachgutachter*innen ihre Meinungen und Einschätzungen zu den Anträgen abgibt. Dies hat zahlreiche Vorteile, z. B. Unabhängigkeit der Begutachtung von der Wahl eines oder zweier Fachgutachter*innen, größere Objektivität, bessere Abdeckung der verschiedenen Beurteilungskriterien, ein besserer Selbstkorrekturmechanismus zur Vermeidung falscher Einschätzungen und bessere finale Beurteilungen. Dies trägt zur Qualität des Begutachtungsergebnisses bei.

¹⁷ Recht viele Befragte führten an, dass Gewicht und Inhalt der verschiedenen Kriterien weiterer Klarstellung bedürften. Da dieses Thema nicht speziell den Peer Circle betrifft, wird es in diesem Bericht nicht vertieft.

prozesses bei. Im Gegensatz dazu gibt es bei der klassischen Begutachtung häufig nur zwei Perspektiven: die des externen Gutachters bzw. der externen Gutachterin und die des Ausschussmitglieds.

Die meisten Ausschussmitglieder finden die PC-Ergebnisse für ihre Empfehlung an den Ausschuss hilfreich, und die PC-Begutachtungen scheinen nicht zu mehr Fragen und längeren Diskussionen im Ausschuss zu führen – was ein gutes Ergebnis ist.

Unter den PC-Mitgliedern gibt es auch relativ unerfahrene Wissenschaftler*innen. Die meisten Befragten empfanden diese Heterogenität als hilfreich und nur ein einziges Ausschussmitglied war der Ansicht, nur erfahrene Fachgutachter*innen könnten sinnvolle Beiträge leisten.

Zusammengefasst kann die erste Frage bejaht werden: Die PC-Gutachten weisen die erforderliche Qualität auf.

2. Früher Konsens

Die Möglichkeit einer vorzeitigen Angleichung der Meinungen trat in dem Experiment nicht oder kaum ein. Es kam zu einer Konsensbildung, aber nicht zu schnell und nicht ohne die Möglichkeit zur Äußerung abweichender Meinungen – was regelmäßig der Fall war. Die Interaktion in der Gruppe und die Möglichkeit, die Stellungnahmen und Kommentare der anderen PC-Mitglieder zu lesen, wurde als positiv, hilfreich und zum Nachdenken anregend empfunden.

3. Akzeptanz

Die Teilnehmer*innen des Peer Circle-Experiments wurden nicht zufällig ausgewählt, sondern kommen aus dem AvH-Netzwerk. Ihre Meinung zum Peer Circle kann daher nicht als repräsentativ angesehen werden – was aufgrund der geringen Größe der Stichprobe ohnehin ein Problem ist. Auf der anderen Seite kommen sie aus unterschiedlichen Forschungsgebieten, sodass eine ausreichende fachliche Vielfalt gegeben ist.

Innerhalb dieser Einschränkungen wurde der Peer Circle insgesamt begrüßt und als guter Ansatz angesehen. Nur drei von 28 Fachgutachter*innen zogen die klassische Review-Methode vor. Diese drei gehörten dem PC Neuere Geschichte an, der insgesamt zehn Fachgutachter*innen umfasste.

Es gab verschiedene Beweggründe für die Teilnahme am Peer Circle, wie Neugier auf das neue Modell zur Begutachtung von Anträgen und das Bedürfnis, durch die Teilnahme an der Begutachtung der Gemeinschaft einen Dienst zu erweisen. Außerdem fanden die PC-Mitglieder es sehr hilfreich, zu sehen, wie andere Fachgutachter*innen arbeiten, was bei der Verbesserung der Begutachtungskompetenzen hilft. Zudem lernt man durch die Sichtung einer Reihe von Anträgen (statt nur eines Antrags wie im klassischen Begutachtungsverfahren), was gute und schlechte Bewerbungen ausmacht – was bei der Verbesserung eigener Anträge hilft.

4. Effizienz

Arbeits- und Zeitaufwand: Bei der Frage, ob der Peer Circle Zeit spart, sind sich die Fachgutachter*innen uneinig. Der zeitliche Schreibaufwand ist geringer, da die Texte kürzer und informeller sind und keine umfassende Argumentation, die in eine Gesamtschlussfolgerung mündet, entwickelt werden muss. Und bei Zustimmung zu Kommentaren / Stellungnahmen anderer Mitglieder ist keine erneute Ausformulierung erforderlich. Jedoch ist der Leseaufwand höher als bei einem Einzelgutachten und dieser wird nicht (immer) durch den geringeren Schreibaufwand aufgewogen.

Die überwiegende Mehrheit ist aber der Ansicht, dass das PC-Verfahren *pro Antrag* weniger Zeit in Anspruch nimmt und dass der Zeitaufwand für den Peer Circle angemessen ist. Die Protokolldateien zeigen, dass es bei der Online-Zeit große Unterschiede zwischen den Fachgutachter*innen gibt.

Unabhängig davon, wie viel Zeit die einzelnen Peer Circle-Mitglieder aufgewandt haben, sollte das PC-Konzept auch aus Sicht der Wissenschaftsgemeinde evaluiert werden. Und für die Gemeinde spart der PC Zeit, da wesentlich weniger Fachgutachter*innen beteiligt sind als beim klassischen Ansatz.

Die Ausschussmitglieder sind unterschiedlicher Meinung. Die meisten geben an, der Zeitaufwand für ihre Aufgaben sei Dank des PC-Inputs geringer, beispielweise weil sie nicht alle Anträge lesen müssen. Bei Anträgen, zu denen im PC überzeugende Einstimmigkeit herrscht, können die Ausschussmitglieder diese Schlussfolgerung übernehmen und sich den anderen Anträgen zuwenden. Andere Ausschussmitglieder sehen dies anders und sagen, sie müssten wie beim klassischen Ansatz alle Anträge lesen. Manche von ihnen haben bei allen oder mehreren Anträgen eine Gutachterrolle übernommen.

Die AvH-Beschäftigten benötigen für den Peer Circle wesentlich weniger Zeit für die Einladung von Fachgutachter*innen als für den klassischen Peer Review. Auf der anderen Seite dauerte die Vorbereitung der Materialien wesentlich länger, da sie nicht in den Standardprozess integriert war. Dies sollte bei den nächsten Versionen der Plattform behoben werden.

Der PC erfordert Koordinierung und Moderation, was eine neue Aufgabe für die AvH-Beschäftigten ist. Diese Aufgabe ist für die Effizienz und die Effektivität des Begutachtungsprozesses entscheidend und sie wurde geschätzt und war effektiv. Wenn der PC im größeren Umfang umgesetzt wird, wird die Koordinierung der Peer Circle möglicherweise für mehr AvH-Beschäftigte zur Hauptaufgabe.

Insgesamt spart der PC der Forschungsgemeinde also Zeit und die PC-Mitglieder halten den Zeitaufwand für angemessen. Für die AvH-Beschäftigten spart er bei der Suche nach Fachgutachter*innen Zeit, aber es wird mehr Zeit für die neue Aufgabe der Moderation und Koordinierung des Peer Circle benötigt. Die Ausschussmitglieder vertreten unterschiedliche Meinungen.

Weitere Effizienzaspekte: Auch wenn sich die Online-Plattform nicht uneingeschränkt für die Aufgaben des Peer Circle eignet, gaben die meisten Teilnehmer*innen an, dass sie vernünftig funktioniere und ihre Nutzung nicht schwer zu erlernen sei. Die in das System hochgeladenen

Dokumente wurden geschätzt, allerdings ließen sich die Navigationsfunktionen verbessern. Die PC-Mitglieder hielten die auf der Plattform bereitgestellten Materialien für besser organisiert als im klassischen Peer Review-Verfahren.

Der Peer Circle war für alle Mitglieder neu. Es wäre gut, wenn die PC-Mitglieder ein paar Jahre lang tätig wären, um Erfahrung zu sammeln.

Einige PC-Mitglieder führten an, dass die aktuelle (Zeit-)Planung der Peer Circle nicht optimal und nicht effizient sei und die Interaktion behindere. Durch eine strukturiertere Planung lässt sich möglicherweise die Effizienz verbessern und mehr Zeit einsparen. Eine mögliche Alternative wäre die Einteilung in z. B. drei kürzere Phasen mit eigenen Fristen: eine anfängliche Begutachtungsphase, eine erste Kommentarphase und eine zweite Kommentarphase. Mehrere PC-Mitglieder fanden es schwierig, die Begutachtungsaufgaben in ihre übrigen Tätigkeiten zu integrieren und schlugen vor, den Peer Circle länger im Voraus anzusetzen.

Die Übertragung der PC-Ergebnisse in einen Vorschlag für den Ausschuss könnte effizienter werden, wenn sich die Ausschussmitglieder stärker auf die Peer Circle-Ergebnisse verlassen würden, was einige, aber nicht alle Ausschussmitglieder bereits tun. Es würde helfen, wenn die Peer Circle eine Rangfolge erstellen würden, koordiniert von den Ausschussmitgliedern. Die Rollen der Ausschussmitglieder müssen wohl überdacht werden, wenn der PC im größeren Maßstab umgesetzt wird. Dies gilt, wie bereits gesagt, auch für die AvH-Beschäftigten.

Schließlich besteht beim klassischen Begutachtungsverfahren das Problem, dass regelmäßig Gutachten zu spät eingereicht werden, was den Entscheidungsfindungsprozess behindert. Verspätete Gutachten führen dazu, dass die Anträge um vier Monate in die nächste Runde verschoben werden. Beim PC tritt dieses Problem nicht auf, da – auch durch die Koordinierungstätigkeit der AvH-Beschäftigten – alles fristgerecht erledigt wird.

5. Werden die Besten ausgewählt?

Laut mehreren Ausschussmitgliedern wäre das Ergebnis des PC-Auswahlprozesses nicht anders gewesen, wenn der klassische Ansatz verwendet worden wäre. Aber wurden die besten Antragsteller*innen ausgewählt? Wir haben uns mithilfe von bibliometrischen Kennzahlen für Anorganische Chemie (PC) und Festkörperchemie (Kontrollgruppe) mit dieser Frage befasst, da in diesen beiden Bereichen Zeitschriftenveröffentlichungen und führende Zeitschriften eine zentrale Rolle bei der wissenschaftlichen Kommunikation spielen. Die Analyse zeigt, dass in beiden Bereichen die allgemeine bibliometrische Leistung der geförderten Antragsteller*innen nicht höher ist als die der abgelehnten, was darauf hindeutet, dass die Veröffentlichungen, die Zitierungen und der Zeitschriften-Impact bei der Entscheidungsfindung weniger wichtig sind als anhand von Beobachtungen aus Gutachtenberichten und Ausschusssitzungen angenommen. Der Ansatz zur Beantwortung dieser Frage ist hilfreich, aber es müssen mehr Qualitätsdimensionen sowie ein größerer Datensatz einbezogen werden.

6. Geschlechtergleichgewicht

Es wird häufig argumentiert, dass sich die Art und Weise, wie Begutachtungs- und Auswahlprozesse organisiert sind, auf das Ergebnis auswirkt. Würde der PC das Geschlechtergleichgewicht in den Ergebnissen des Auswahlverfahrens beeinflussen? Ein Vergleich der Erfolgsquoten der klassischen Methode und des Peer Circle (in den acht Fachgebieten aus dieser Studie) zeigt, dass die Erfolgsquoten von Männern und Frauen in beiden Gruppen gleich sind. Auf Ebene der einzelnen Forschungsgebiete gibt es bei den Ergebnissen erhebliche Unterschiede zwischen den Gebieten und zwischen den Jahren. Ob dies eine Verzerrung oder leistungsabhängige Geschlechterunterschiede darstellt, kann noch nicht beantwortet werden und würde mehr Daten und eine weitere Analyse erfordern. Zum jetzigen Zeitpunkt kann man schlussfolgern, dass der Peer Circle sich neutral auf die Chancen von Frauen auswirkt, dies gilt es jedoch weiter zu untersuchen.

10. Empfehlungen

Aus dieser Studie ergeben sich die folgenden Empfehlungen:

- Fortführung des Peer Circle für Förderprogramme, bei denen der Schwerpunkt mehr auf den Antragsteller*innen als auf dem Projekt liegt.
- Änderung der (Zeit-)Planung des PC, so dass es feste Zeiträume für die verschiedenen Aufgaben mit eigenen Fristen gibt. Beispielsweise eine Phase für die Begutachtungen durch die Erstgutachter*innen, eine Frage- und Kommentierphase und eine letzte Phase für Diskussionen und möglicherweise die Erstellung einer Rangfolge für den Ausschuss.
- In diesem Zusammenhang Überdenken der Aufgaben der Ausschussmitglieder.
- Umsetzung eines ähnlichen Benotungssystems im PC, wie es beim klassischen Ansatz verwendet wird.
- Planung des Peer Circle länger im Voraus, so dass sich die spezifischen Tätigkeiten besser in die Terminplanung der PC-Mitglieder integrieren lassen.
- Entwicklung eines Verfahrensprotokolls, aus dem hervorgeht, was in welcher Phase des PC zu erledigen ist. Dies würde auch zu einem gleichmäßigeren (und bei manchen zu einem höheren) Aktivitätsniveau führen.
- Da die Koordination für den PC von entscheidender Bedeutung ist, Entwicklung eines entsprechenden Verfahrens, da die Koordination im Experiment in jedem Peer Circle unterschiedlich gehandhabt wurde.
- Die neue Plattform sollte auf Englisch sein. Derzeit läuft die Kommunikation im Peer Circle überwiegend auf Deutsch ab, was eine Hürde für die Teilnahme von nicht deutschsprachigen AvH-Preisträger*innen und anderen potenziellen Teilnehmenden darstellt.
- Künftige Experimente könnten in größerem Maßstab durchgeführt werden, da dies eine umfassendere Evaluation ermöglicht.