publications

MDPI

Article

Experimenting with Grant Peer Review: A Mixed Methods Case Study of the Effects on Time Use and the Quality of Reviewing

Peter van den Besselaar [1,2,*] and Charlie Mom [2]

[1] Department of Organization Sciences, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands
[2] TMC Research, 1098 AN Amsterdam, The Netherlands; charlie@teresamom.com
* Correspondence: p.a.a.vanden.besselaar@vu.nl or peter@teresamom.com

Abstract

Rejection of review invitations due to time constraints is putting pressure on the peer review system, showing that less time-consuming ways of reviewing are needed. This paper presents the results of a field experiment with a new format for grant peer review and answers the question of whether this new format is less time-consuming while still providing high-quality reviews. In the new approach, the Peer Circle (PC), a team of reviewers collectively evaluates several grant applications. The PC was applied to four fields and compared with four similar fields using conventional peer review. Qualitative and quantitative methods have been used to analyze heterogeneous data such as interviews with and a survey of the peer reviewers; text analysis of the review reports; and statistical analysis of bibliometric applicant data. The comparison suggests that the PC saves time and enlarges the reviewer population considerably. Most reviewers felt that the quality of the PC evaluations was at least as good as that of the conventional evaluations, if not better. Given these findings, the experiment is now continued on a much larger scale. Apart from that, the theoretical implication is that the way of organizing peer review has an important effect on the functioning of the system.

Keywords: research grants; peer review; new models of peer review; peer circle

1. Introduction

Peer review is the dominant evaluation system in science, a procedure where scientists and their products are assessed by peers, that is by other specialists in their field. It is perceived as the gold standard for academic evaluation of paper submissions, academic job applications and research grant applications. In the context of research funding—the topic of this paper—it is estimated that up to 95% of all grant applications go through peer review. The evaluation process requires several (at least two but often more) independent reviews, which is increasingly a problem because the number of declined invitations to participate in peer review is high and rising, which undermines the viability of the peer review system itself. Reviewing is labor intensive, and the main reason why researchers often reject a review invitation is that they do not have the time for it (Publons, 2019a, 2019b; Belém de Oliveira Neto, 2024; Oxford University Press, 2024; Adam, 2025), which has been an issue reported at least since the middle of the 2000s (Tite & Schroter, 2007). This is the main problem we address in this paper.

Independence in the context of peer review means two different things. One is that reviewers should be independent from the applicant to avoid conflicts of interest. This is needed, as peer reviewers are inclined to rate those they are related to higher than unrelated

Check for updates

Academic Editor: Andrew Kirby

Received: 25 January 2026
Revised: 11 May 2026
Accepted: 13 May 2026
Published: 27 May 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

applicants both in article peer review (Teplitskiy et al., 2018) and in grant peer review (Jerrim & de Vries, 2023; Mom & Van den Besselaar, 2022). The other meaning of independence is that the different reviewers of a grant application should review independently of each other, which is the common practice in grant peer review (Kaltenbrunner et al., 2021; Publons, 2019b). This independence is important as it enables a pluralistic assessment of the quality of the research proposal (Abramo & D'Angelo, 2025; Baccini & Re, 2025), avoiding processes like groupthink leading to premature consensus. Research indeed shows that independent reviewers rate the same grant application rather differently (Cicchetti, 1991; Jerrim & de Vries, 2023), and that needs to be handled in the next step of the selection process where a panel discusses the reviews to come to a ranking (Oxley, 2025). The ranking then either determines who gets the grant or is used as advice by the decision-making committee.

Despite its general use, peer review of grant proposals faces a number of problems. A main problem is mentioned above: the growing lack of competent reviewers due to time constraints. Other problems are the lack of (predictive) validity and reliability of the review outcomes. These problems require developing other, less labor-intensive ways of evaluating grant applications, and these alternative formats should lead to reviews that are at least as good as the conventional reviews but preferably better. Several quality aspects play a role in such a comparison, such as (i) the level of scope and detail of the reviews; (ii) the possible occurrence of premature consensus which would hinder a pluralistic and critical assessment of applications; (iii) the validity and reliability of the review; and (iv) whether one can expect that an alternative model will be accepted by the scientific community.

In this paper we report the findings of a case study of an alternative form of peer review, the Peer Circle (PC) developed by one of the large German research funders. In this exploration of the PC, the following two research questions are addressed:

- Does the PC contribute to solving the problem of the shortage of peer reviewers?
- Is the quality of the PC reviews at least as good as the conventional reviews?

In this multi-method case study, data were collected from various sources and analyzed qualitatively (content analysis of interviews, and logfiles of the PC platform) and quantitatively (administrative data, survey, bibliometric data, review texts, application texts). This enables combining the opinions of the reviewers about the PC with more objective results from data that reflect the behavior of the PC reviewers and the outcomes of the assessment process. Combining the findings of the different analyses based on those heterogeneous data leads to more robust findings than would be possible with only a single data source. The experiment with the PC was done in four research fields, and these are compared with four other fields that are similar in terms of discipline and disciplinary culture. The study provides an affirmative answer to both research questions.

The remainder of this paper starts to discuss the problems with peer review (Section 2), followed by a discussion of some proposed alternative forms of peer review and alternatives for peer review (Section 3). In Section 4, the PC is described in some detail, followed in Section 5 by a description of the data and methods of the study. Section 6 presents the findings, and the paper ends with conclusions and a discussion (Section 7).

2. Peer Review and Its Problems

Scientific peer review is the evaluation of a scientist, a scientific product (articles), a job application, or a grant application[1] by one or more other scientists coming from the same or a similar research field. This paper focuses on peer review of grant applications, but where useful, also refers to research on article peer review. Peer review of grant applications is the first step in the process of distributing research funding. It generally starts with independent reviewers evaluating the grant application in terms of the criteria formulated

by the funding organization, and these criteria may be different depending on the type of funding or funder (Hug & Aeschbach, 2020; Mocanu et al., 2024). In the next step, the reviews are discussed in a (disciplinary) panel, and reviewers may be members of the panel. The panel comes to a final score for each of the applications that belong to the domain of the panel. The assessment by the panel can happen in one or two rounds and can include an interview with applicants. The panel produces a ranking, which can be directly decisive for who gets funding or provides advice to the decision-makers.

The reviewers should be independent from the reviewed person and their work. In the conventional form of peer review, independence between the reviewers is also considered important, which means that a reviewer does not know who else is reviewing nor are the reviewers aware of the content of the other reviews. This enables a pluriform assessment (Abramo & D'Angelo, 2025) and may also reduce bias. To ensure the possibility of pluralism in reviews, using two reviewers is the bare minimum. To find two reviewers, multiple potential reviewers must be invited because scientists often reject the review invitation due to various reasons. The main issue is time constraints: Scientists report that they get too many invitations to review and that it takes too much time (Publons, 2019b).

In contrast to article peer review, peer review of grant applications is seldom double-blind, as it is needed to assess whether the applicant would be able to carry out the proposed project. Applicants generally do not have the possibility to revise the application based on the reviews, although the applicant may get the opportunity to respond to the reviews.

However, peer review is not undisputed. At the latest since the early 1980s, studies critical of peer review were published (J. R. Cole & Cole, 1981; S. Cole et al., 1981; Chubin & Hackett, 1990) and the debate continues (e.g., Marsh et al., 2008; for a recent overview: Aczel et al., 2025). Some issues apply for both grant and publication peer review: (i) problems in engaging enough (qualified) reviewers; (ii) bias for and against persons, groups, countries or regions, theories, methods and ideas; (iii) a declining quality of reviews, due to the increasing theoretical and methodological complexity of papers and grant proposals; and (iv) the ongoing debate about low (inter-rater) reliability and validity (Aczel et al., 2025). To have reliable results, one would need quite a few reviewers in the range between five and seven (Snell, 2015; Mayo et al., 2006; Marsh et al., 2008). And validity requires taking into account a variety of quality dimensions such as accuracy, clarity, consistency, novelty, replicability, and thoroughness (Smith et al., 2025).

There are also differences between grant peer review and article peer review. Uncertainty in evaluating research grant proposals is much higher, as research plans are evaluated and not research outcomes (knowledge claims in papers). When reviewing a paper one should be led by the content of the paper, not by the reputation or past performance of the author. When grant applications are evaluated the expected output cannot be assessed yet, and therefore reviewers must also look at past performance to assess whether the applicant is able (has the skills) to successfully carry out the proposed research project. Another problem is that success rates of funding schemes are often low and declining, leading to increased competition for scarcer resources and likely resulting in more false negatives: Rejection of excellent applications because of insufficient money. With low acceptance rates, the costs of the selection process become large compared to the amount of funding distributed. And the costs are even higher if one includes the time to write the rejected applications (Schweiger et al., 2024).

In the case we study in this paper, the rejection rate of review invitations is 60%. Exacerbating this shortage of reviewers, quite some reviews that were promised did not arrive or arrived too late for the selection process. Consequently, 100 of the 235 applications were only reviewed once, instead of the mandated minimum of two. These kinds of problems necessitate innovation in the peer review system. Can one devise alternative

models for peer review that cost less time and money, and can this be done without compromising on the quality of the assessments, or given the criticism of conventional peer review, while improving the quality? As the awareness of these problems has become larger, alternative peer review formats have been proposed, as well as alternatives to peer review.

3. Alternative Formats for Grant Selection and Peer Review

To solve the problems, models were proposed for selecting grant applications without peer review. Vaesen and Katzav (2017) proposed to divide available research money over all qualified researchers without peer review and discussed the possible consequences of such a model. They found that "researchers would, on average, maintain current PhD student and postdoc employment levels and still have at their disposal a moderate (the U.K.) to considerable (the Netherlands and the U.S.) budget for travel and equipment." However, some kind of review may slip back in as there needs to be a (peer review-based) selection of who counts as a 'qualified researcher'. Bollen et al. (2014) proposed a highly decentralized funding model, allocating a fixed amount of public funding to all researchers, but each is in turn required to redistribute some fraction of the total they received to other researchers, making all scientists both researchers and reviewers. Bollen and colleagues expect that such a funding system may result in significant changes in scholarly communication and collaboration. This is because everyone needs to attract the interest of colleagues and convince them of the importance of one's research agenda as the colleagues are potential donors. One risk of the model is that the allocation of funding may depend more on popularity than on quality and relevance. Another unwanted effect of both models is a reduced capacity for thematic funding, as that requires a review of the thematic content.

Getting rid of competitive funding whatsoever may not be a very attractive solution for the problems with peer review, as it neglects any positive effects competition for research grants may have on the science system. Writing competitive grant applications may improve the quality of applications and the quality of the related output, an advantage that should not be neglected (Schweiger et al., 2024). Neither proposal was taken up by research funders, in contrast to the lottery model (Adam, 2019; Roumbanis, 2019, 2024). As full lotteries may create noise by missing top applications and by awarding low-quality applications, a number of hybrid lottery models (Feliciani et al., 2024; Horbach et al., 2022; Philipps, 2022) have been proposed that combine a lottery with peer review: For example peer review is used to select the most excellent applications and to reject those of insufficient quality, and a lottery then selects from the 'gray middle range' who else gets funding. Experiments with lotteries have been done with positive evaluations, and the acceptance of lotteries seems to increase over time—especially among early-career researchers (Barlösius & Philipps, 2022; Heger, 2025a, 2025b).

Additionally, alternative forms of peer review have been suggested. To improve reliability and to counteract (gender) bias, structured reviews with mandatory checklists were suggested, in line with findings in psychology (Kahneman et al., 2021). Also using anonymized grant applications would contribute to the elimination of bias (Witteman et al., 2019). To solve the problem of reviewer availability it was proposed to use AI to support reviewing. Research funders have begun publishing guidelines for this, allowing rather restricted use only (e.g., ERC, 2026). Studies started to explore the use of AI in peer review (Yang et al., 2026), but we could not find research on AI in grant peer review. Finally, because reviewing the project proposal takes most of the time, it was suggested to review only the project abstract plus the applicant's CV (Simsek et al., 2024).

Remuneration has been proposed as a solution to attract more reviewers. For article reviews it seems rarely used, probably because of the costs involved which would increase APCs considerably. Small experiments done by some journals showed small to moderate

effects on reviewer availability and time to completion (Siemroth, 2024; Gorelick & Clark, 2025; Cotton et al., 2025). But an experiment by Squazzoni et al. (2013) suggests that remuneration for both article and grant peer review has mainly negative effects and undermines the idea of peer review as a community duty. However, remuneration is used substantially in grant peer review, predominantly for reviewers who also act as panel members (Publons, 2019b). But we could not find any studies on its effectiveness in terms of the availability of reviewers and the quality of reviews.

Distributed peer review, a system where applicants act as reviewers of other applicants is currently in use, e.g., for the UKRI Metascience AI early career fellowships. This model was chosen because current AI is a fast-changing field, making it difficult to find peer reviewers, which then causes additional delays. Therefore, UKRI believes that applicants are the best fit for reviewing grant applications (UKRI, 2025). This model solves the peer reviewer shortage as reviewing is required for applicants. However, the model has some disadvantages. Reviewers are not independent anymore as they have a stake in the outcome: grading competitors low increases one's own chances. To counteract that, UKRI rewards high inter-rater reliability: The more an applicant/reviewer's review resembles those of other reviewers, the more additional credit she gets in the final grant decision. This solution unfortunately creates an incentive for conservative reviewing. Applicants may also fear insufficient confidentiality and may consequently submit their application to another funding scheme (Brainard, 2025) if available.

In this paper we focus on another alternative form of peer review, the Peer Circle (PC), which has been used since 2022 in real-life experiments. The PC differs from conventional peer review in that (i) it organizes the review as a collaborative group process reviewing a set of applications, implying that the contributions are not independent; and (ii) as the review is a group process, the population of eligible peers could be firmly extended by allowing group members to review only part of the proposals that are within their core expertise. While keeping peer review at the center of the selection process.

4. The Peer Circle

This paper presents a study of a field experiment of the Alexander von Humboldt Foundation (AvH)[2], a main research funder for international grants in Germany, with another alternative to conventional peer review: the Peer Circle. The experiment was carried out between March and November 2022. The grants of the AvH enable foreign researchers to come to Germany and German researchers to go abroad. The HFST[3] grant, which is the case in this study, is meant for early-career researchers to do a postdoc in Germany and for more advanced researchers to do a research project up to two years at a German university.

The conventional peer review procedure of the AvH Foundation includes two independent reviewers for each of the applications, and these reviewers are meant to be real experts on the research topic. The review should cover: (1) the quality of the CV, (2) the quality of the core publications, (3) the quality of the proposed project, and importantly (4) the applicant's career prospects. After the reviews have been received, the field representatives give advice to the selection committee (consisting of all field representatives) based on the reviews and on their own assessment. The field representative classifies the applications into three groups: 'reject without committee discussion' (R-case), a limited number of 'grant without discussion in the committee' (G-case), and a group that should be discussed in the selection committee before deciding on those (D-case). All reviews and all applications are made available to all committee members, and in the committee meeting some of the G-cases and the R-cases may move to the D-case category on request of one of the other

committee members. After the discussions on the D-cases, there is a secret grading and the outcome of that determines which D-case applications will be funded.

The new Peer Circle procedure is a group-based online evaluation. A group of about 7 reviewers are together responsible for reviewing about 11 applications per round, using a web-based platform where the PC members provide and discuss their assessments of the applications. In practice this means that one or two group members who are closest to the application write a full review, but not as extensive as in the conventional procedure, covering the four dimensions mentioned above, and the other members review those parts of the applications they are familiar with, given their expertise and experience. So, someone could focus only on the statistical part of the application and assess whether the proposed analytical strategy is adequate. The reviewers are expected to comment on one another's (partial) reviews and comments, thereby creating a discussion. In a few cases, the PC lacked specific expertise to assess an application, and in those situations an additional reviewer was added to the PC for that individual application. In the experiment, the PC handled two rounds in one year, but the normal frequency would be three rounds per year.

One PC round lasts a few months without face-to-face interaction which gives PC members flexibility in doing the reviews. But it was suggested to concentrate activities in two periods of two weeks to facilitate online interaction. The conventional reviewers are generally experienced professors; the PC also includes less experienced researchers often in their early careers. Anonymity within the PCs was maintained to offer less experienced members a certain degree of security so that they can express their opinions (in writing), even if these conflict with the views of well-established PC members. After the PC finishes the review, the results (which are substantive answers on the review questions, but no grading or ranking) are handed to the field representative, who uses this for his/her advice in the committee meeting.

In the experimental phase of the PC, 'G-case' advice could not be given, but only 'R-case' or 'D-case' advice. Like in the conventional approach, the field representative presents the applications and his/her advice to the committee, which are then discussed and graded in the committee leading to the funding decision.[4]

Summarizing, the main differences between the two procedures are firstly that in the conventional remote review model the reviewers do not see any other application or review, and they do not discuss the proposals or reviews with each other. The discussion about the applications and reviews takes place in the committee only. In contrast, in the PC the reviewers discuss the applications and reviews before this information goes to the committee, implying that the PC reviews are more comparative. Secondly, the PC is organized via anonymous online interaction over a longer period, so it is a flexible group effort: One can read the contributions of other reviewers and post one's own contributions when convenient within the period given. Thirdly, by allowing for partial reviews, a larger set of potential reviewers is available, and the issue about partial familiarity with the applications is explicitly addressed. Fourthly, the communication is written; it requires more precision than if it were oral like in panels or committees. Figure 1 summarizes the differences.

How does the PC differ from the common panel review (Oxley, 2025)? The main difference is at the level of the reviews: In the PC, the reviews are not done independently before the panel meeting but are done interactively and based on various inputs from different PC members, who can provide a full review or a partial one focusing on a few elements only. As will be shown in Section 6, ex ante prepared reviews are of a different nature than collaboratively prepared reviews. And different from panels, PC members do not need to be experienced specialists who can and have to cover all aspects of an application.

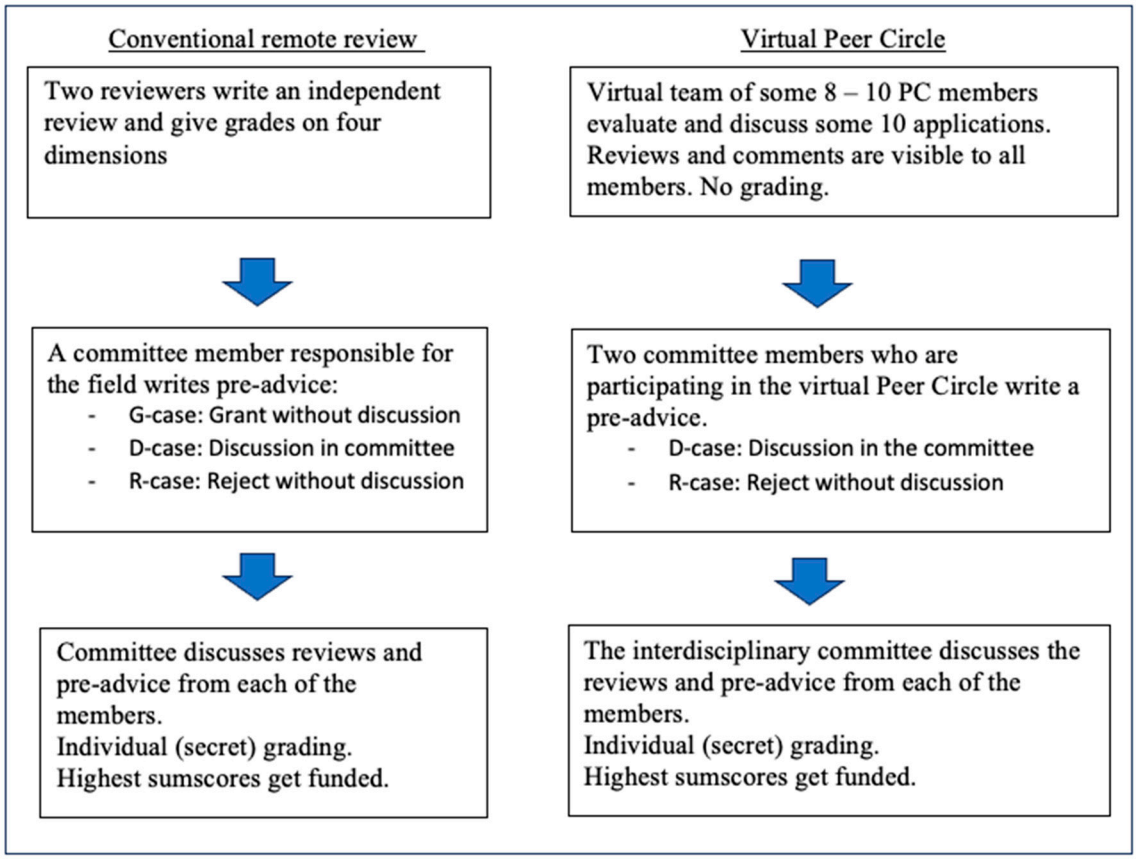


Figure 1. The two procedures at the AvH Foundation: conventional peer review versus Peer Circle.

5. Study Design, Data and Methods

For this study, the PC was implemented in 2022 in four fields during the summer and autumn rounds. By having the same PC do the reviews in two rounds, one can account for learning effects. To enable some generalization of the results, the fields were selected from different disciplines: natural sciences, technical sciences, life sciences and humanities. To find out whether the PC potentially can solve the problems of conventional peer review, the PC members were interviewed and surveyed about their experiences and opinions. As most of them had earlier experiences with conventional grant peer review, we also asked them in the interviews to reflect on the differences between the two models of peer review. This was accompanied by a quantitative comparison of the time investment required in the two models of peer review. An important question to be answered was about the quality of the PC reviews in comparison to conventional reviews. We addressed this in the interviews, but we also compared the texts of the reviews. For this we used the PC reviews in the four fields (inorganic chemistry, material science, zoology, and modern history) and the conventional reviews in four fields with a comparable disciplinary culture (solid state chemistry, materials engineering, botany, and ancient history) in 2022. We also included in that comparison the conventional reviews in the eight fields in 2021, to avoid that the findings are influenced by annual fluctuations (Figure 2).

Table 1a shows the number of applicants involved in this study and Table 1b shows the number of PC members. The number of applicants with a conventional peer review adds up to 235, which would need 470 peer reviewers. However, about 370 peer reviewers were involved (not in Table 1), and 100 of the 235 applicants were reviewed by only one expert, illustrating the problem of finding enough reviewers. In some cases, the committee member had the proper specialization and was appointed as the second reviewer. If there are not enough reviewers, the committee can choose to decide based on only a single review.

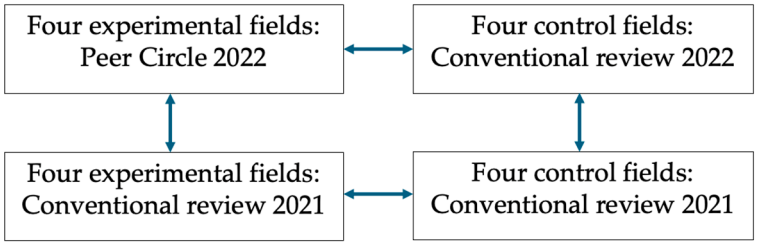


Figure 2. The comparative approach.

Table 1. Number of applicants and of reviewers.

(a) Number of Applicants		
Experimental fields	2021 conventional	2022 Peer Circle
Inorganic Chemistry	29	20
Materials Science	13	19
Zoology (biodiversity)	24	30
Modern and Contemporary History	32	20
Total	98	89
Control fields	2021 conventional	2022 conventional
Solid State Chemistry	21	14
Materials Engineering	16	21
Plant Science	15	19
Ancient History	17	15
Total	69	69
(b) Number of PC Reviewers		
Experimental fields	2022 PC members	2022 if conventional
Inorganic Chemistry	5	40
Materials Science	9	38
Zoology (biodiversity)	6	60
Modern and Contemporary History	9	40
Total	29	178

The following data were collected about the PC:

- We had access to the *assessment and administrative* data. This covered the committee scores the applicants received and the rankings and decisions, as well as the personal characteristics such as age, nationality, gender, residence, institutional affiliation, planned host organization (university or public research organization) in Germany, and the current position.
- PC members, committee members and involved AvH staff were *interviewed* about their (self-reported) PC behavior, about the differences between the PC and conventional peer review and how they value these, and about their opinions of the PC. In total 56 interviews were conducted. The interviews were semi-structured following a list of topics derived from the research questions, with as much openness as possible to allow the interviewees to address everything they found important. Interviews were done after the summer round and after the autumn round.
- PC members were twice *surveyed* for similar information as was collected in the interviews, to increase the validity and reliability of the measured opinions and behavior.

This self-reported evidence will be combined with more objective observational data, to increase the reliability and robustness of the analysis:

- *Logfiles* provide information about the online activities on the PC digital platform: Frequency of logins, total time online, and total time active online. The logfiles also provided the frequency and dates of review contributions and of the comments. As the interviews and survey also addressed time used, the self-reported and observational data can be compared, which increases reliability.
- Committee meetings were *observed* to measure (i) the duration of the presentation and discussion devoted to each application; (ii) the number of participants in the discussion of each application; and (iii) differences between handling the PC-reviewed applications and the conventionally reviewed applications.
- The conventional *review texts* were used to extract the text parts and the review scores for each of the review dimensions: the applicant's future perspectives and the quality of the CV, of the core publications, and of the proposed project. The PC *reviews and comment texts* were extracted and categorized in the same four dimensions. These texts were also used to analyze the styles of the reviews. This provides information about the nature of the review process, which can be compared with what the PC members reported in the interviews and surveys.
- To answer the question of whether the Peer Circle is at least as good as conventional peer review for identifying the best applicants, we collected *bibliometric data* measuring the output and impact of applicants. For the applicants in the chemistry panels, publications were searched in Scopus, and some bibliometric indicators were retrieved from SciVal.

To compare the PC with the conventional way of reviewing at the AvH, a mixed-methods approach has been used. The opinions of the PC members about relevant aspects of the PC procedure and results were extracted from the interviews. The interview topic list was used to index the interview text. The topic list contained among others the quality dimensions distinguished in this paper, the various aspects of the time needed for doing the reviews, and whether the PC would be a better and acceptable alternative for conventional peer review. The survey outcomes were summarized using descriptive statistics.

The linguistic data to describe and compare the review reports were produced with the Linguistic Inquiry and Word Count (LIWC) tool[5] (Tauzik & Pennebaker, 2010; Pennebaker, 2011). LIWC provides descriptives about the texts, such as *word count* and average *sentence length*, plus a library of word categories reflecting linguistic properties of texts, which can be used to statistically compare review texts (Van den Besselaar et al., 2018). An example is the category *writing style*, which varies between a strongly *analytic style* and a strongly *narrative style*. Text analysis has shown that analytic writing is characterized by a frequent use of articles and prepositions and an infrequent use of pronouns, auxiliary verbs, adverbs, negations and conjunctions. And narrative writing is characterized by an opposite pattern of the use of these function words. Pennebaker et al. (2014) derived the following scale, originally called *CDI* but in LIWC called *Analytic writing*, which can be applied in the analysis of review reports and grant applications (Markowitz, 2019; Van den Besselaar & Mom, 2022):

$$30 + \%\,\text{articles} + \%\,\text{prepositions} - \%\,\text{pronouns} - \%\,\text{auxiliary verbs} - \%\,\text{adverbs} - \%\,\text{negations} - \%\,\text{conjunctions}$$

LIWC contains a large number of predefined word categories, but the user can also create specific word categories. For this study, we created within LIWC targeted categories with terms reflecting the various evaluation dimensions used by the AvH allowing us to measure their relative importance (see Section 6.2 below).

6. Findings

This section starts with the findings about the reviewer shortage and then discusses the quality of the PC reviews compared to the conventional reviews, using six quality dimensions. PC members have articulated their opinions and experiences, and these are compared with the more objective data we collected through (online) observation and text analysis.

6.1. Does the PC Help to Alleviate the Lack of Reviewers

The AvH needs some 5000 reviewers for its funding instruments every year. To solve the reviewer availability problem, the PC aims to increase the population of potential reviewers and to decrease the time needed for reviewing. For conventional review, mainly full professors are invited, who are predominantly older men. In contrast, the PC reviewers are more diverse, with 43% women, and 36% and 32% are in their thirties and forties respectively. The share of non-professors is substantial at 54%. This indicates that the PC indeed recruits from a larger population than is the case in conventional peer review. Are those younger reviewers as active as the more established ones? The log-file analysis confirms this: The correlation between age and the time active on the PC platform is negligible, and the correlation between age and commenting activity on the platform is low ($r = 0.10$) and non-significant ($p = 0.34$).

Does the PC reduce the time needed for reviewing? Several younger PC members had never reviewed grant applications before and could not answer this question. Of those who did have previous experience, most reported in their interviews that the time used for the PC was reasonable, which was supported by the survey where 75% of the respondents agreed with this. Whether the PC saves time compared to conventional peer review shows a more diverse pattern. Those who feel that the PC takes more time point to the larger number of applications that one should read, which takes more time than just reading one as in the conventional peer-review setting. Additionally, one also has to read the reviews and comments from other PC members. On the other hand, those who argue that the PC saves time argue that one does not need to read all applications fully, which is supported by the survey where 80% reports that they only do so with one or a few applications. PC members also report that reading others' comments significantly helps to review applications, as does the fact that the reviews are structured as questions. The possibility to comment on others' contributions enables agreeing or disagreeing without repeating the arguments. Finally, PC members reported that their reviews are written in a much more informal style, whereas the conventional reviews are composed texts with an argumentative style leading to a verdict about the application. This all substantially reduces writing time compared to the conventional review approach, which is the time-consuming part of reviewing.

Several interviewees believe that the PC takes per application less time than a conventional review, but that PC membership overall requires more time. Conventional review requires a reviewer to assess a single application; PC members contribute to the evaluation of a set of applications, which together may take more time. However, as interviewees emphasized, the PC is more efficient for the academic community: In total, significantly fewer reviewers are needed. The four experimental fields with 89 applications included 29 PC members (plus sometimes an additional reviewer for specific applications for which the PC lacked the knowledge), which in the conventional review procedure would have required 178 reviewers (two per application)[6]. This means that 29 PC members did the reviewing that traditionally would have required six times as many reviewers. And if the assessment of the PC members is correct that the PC takes less review time per application, then the total time required for reviewing becomes considerably lower, which is also more

evenly distributed within the community because the PC draws from a more diverse population than the AvH did in its traditional approach. Even if PC members on average spend more time, the total time investment is considerably lower. Implementing the PC can be expected to reduce the pressure on the review system and consequently reduce the need for other measures like paying reviewers.

As discussed above, we combine subjective assessments of the PC members with objective data to arrive at robust conclusions. According to the interviewees, most activities were done online and the logfiles register the amount of time logged in and the amount of time active online. The latter can be used to estimate the time PC members spend on writing the reviews and comments. The former can be interpreted as the maximum time spent on reading applications—maximum because one can be online while doing things unrelated to the PC.

Logfiles can overestimate and underestimate time use. From interviews with the AvH staff we know that not all activity was correctly logged by the system (underestimating) and the logfiles do of course not include offline activities (underestimating), while on the other hand reviewers can be logged in while doing something else (overestimating). The registration of 'hours active' only suffers from underestimation. So, we feel safe when interpreting hours active as hours spent on writing, but total hours active can be only a very noisy indication of total time spent on the PC. The findings from the logfiles are shown in Table 2. In the first (summer) round, on average the PC members spent half of a workday "active online" (Table 2). The time spent was halved in the second round, suggesting a learning effect. The very low minimal scores show that not all PC members have been very active. However, they may also have written offline and then uploaded their contributions. In the interviews with the PC members done after round 1, eight members estimated the time used for the PC between one and eight days, with an average of four working days spent, which matches the average 32 h logged in time (Table 2).

Table 2. Login time *—online platform.

	Hours Logged In		Hours Active	
Experimental Fields	**Round 1**	**Round 2**	**Round 1**	**Round 2**
Average	31.9	14.2	4.2	1.9
Minimum	6.9	0.2	1.1	0.1
Maximum	77.1	49.7	9.4	7.9

* Not including the field representatives, the AvH staff, and three reviewers without any activity. Source: Logfiles.

We observed the second step in the procedure, which is the grading of the applications in the committee. Observing the committees, we could measure the time used for discussing the applications and found that in the case of the conventional reviews, there was more discussion in the committee than for PC-reviewed applications (Figure 3). More of the conventionally reviewed applications were discussed intensively, suggesting that more uncertainty remained after the conventional review process. The discussions in the decision-making committee about the conventionally reviewed applications took more time than PC-reviewed applications, indicating that the PC also saves time in this step in the procedure.

In conclusion, the PC members found the time investment reasonable, and overall, the review process in the PC costs considerably less time and fewer reviewers than the conventional review does. The PC therefore contributes to alleviating the reviewer problem the current funding system is suffering from.

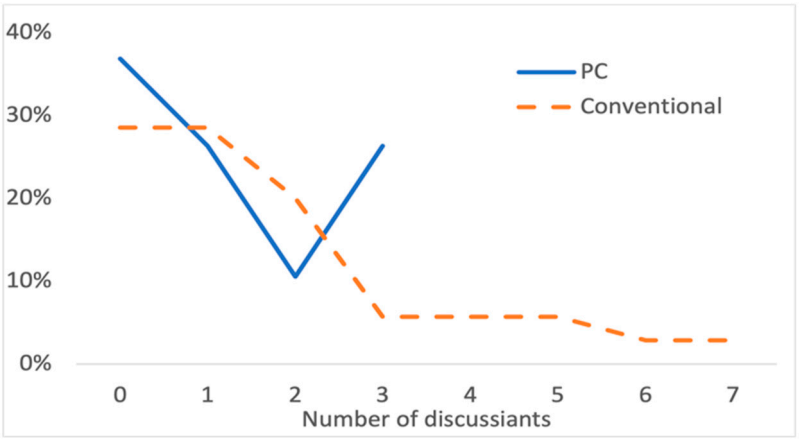


Figure 3. Discussants per application in the committee.

6.2. The Quality of the PC Reviews

Quality is a multidimensional concept, and in this study several quality dimensions are addressed to compare the Peer Circle with the conventional reviews (Table 3).

Table 3. Quality dimensions addressed.

- Review style
- Use of evaluation criteria
- Premature consensus and the interaction between PC members
- Identifying the best applicants
- Perceived overall quality
- Acceptance by the scientific community

Review style

In the interviews, the PC members reported that the PC reviews and comments were more informal than the conventional reviews. An interviewee remarked that "the comments are open and honest", and another described it as "a less curated review, less effort to make a coherent argument that leads to the selected verdict". This is confirmed by a linguistic analysis:

- Conventional reviews have a strong analytical style, scoring on average 97 measured on a scale from 0 to 100. In contrast, the PC reviews score 87, significantly lower ($F(1, 322) = 89.87$, $p < 0.000$), showing that the PC reviews are more narrative.
- Clout measures—also on a 100-point scale—how strongly a text emphasizes authority and standing. PC reviews score significantly lower than the conventional reviews: (46 versus 61; $F(1, 322) = 159.30$, $p < 0.000$).
- Finally, the authenticity score measures whether the review text is written in an honest, humble and vulnerable way. The PC texts score on average 48 (out of 100) on authentic whereas the average for the conventional reviews is 32, a significant difference ($F(1, 322) = 99.14$, $p < 0.000$).

The PC appears to have a more open social dynamic that is more honestly evaluating the grant applications, suggesting a more equal relationship between applicants and reviewers, which can be considered an advantage if one takes the peer aspect of peer review seriously.

Use of evaluation criteria

During the interviews, PC members discussed the use of criteria, arguing that topical specialists tend to focus on the scientific content of the research project, often at the expense

of other important review criteria. It was doubted that technical details—relevant for feedback—should be central in the research grant selection. This is a pertinent issue because the HFST program provides an individual career grant based on several criteria. The proposed project should be valuable, but the applicant should be assessed in the first place in terms of past achievements and future potential. Interviewees felt that PC was the better model to evaluate such applications, as the varied membership is more likely to cover the broader evaluation criteria. A reduced focus on technical aspects was seen as a positive development because it reduces the risk that technicalities become too central in the grant selection process. Technical problems can be solved, and technical weaknesses can be improved on during the research.

Analyzing the review texts can clarify whether the assessment criteria have the same emphasis in the PC as in conventional reviews. We created a LIWC dictionary with terms used in the review texts that relate to the assessment criteria used for the HFST grant: career quality and career prospects; mobility; independence; quality of the (core) publications; school and university performance; excellence; the quality of the project; and the quality of the host organization where the applicant is going to work. Terms were extracted from all review reports, and two researchers independently selected the relevant terms for each of these evaluation criteria. Using LIWC, the frequency of terms belonging to each criterion was calculated for both the PC reviews and the conventional reviews, which have not yet been standardized at the field level.

Table 4 shows that PC assessments contain a significantly higher percentage of non-technical ('common') terms. This corresponds with the interviewees' statements that the PC focuses less on technical aspects and more on the other assessment criteria. It also corresponds with the observation that the conventional reviews contain significantly more project-related terms. And although conventional reviews and the PC reviews have a similar emphasis on performance when combining both performance evaluation criteria, the PC places slightly more emphasis on bibliometrics. Finally, the host institution and the applicant's independence are significantly more discussed in the PC reviews. These findings confirm the somewhat broader use of criteria by the PC compared to the more technical project orientation of the conventional reviews.

Table 4. Emphasis on the various evaluation criteria.

Evaluation Criteria	Peer Circle Review	Conventional Review	One Way Anova
Common (non-technical) words	80.0%	76.0%	$F(1, 322) = 46.03$, $p < 0.0000$
Career (incl. mobility)	0.93%	0.96%	n.s. *
Performance (incl. school and university)	0.60%	0.72%	$F(1, 322) = 5.03$, $p = 0.0257$
Publication performance (bibliometrics)	1.26%	1.17%	n.s.
Proposed project	1.04%	1.24%	$F(1, 322) = 9.14$, $p = 0.0027$
Independence	0.07%	0.02%	$F(1, 322) = 31.52$, $p < 0.0000$
Excellence	1.39%	1.47%	n.s.
Host	0.23%	0.17%	$F(1, 322) = 7.00$, $p = 0.0086$
Final score by the committee	0.55%	0.58%	n.s.

* n.s. non-significant.

Premature consensus and interaction between PC members

PC reviewing is a team activity. This has two aspects: one is reading each other's contributions, and the other is reacting to the contributions of others: in online discussions. The interviews suggest that there was some discussion, slightly more in the (second) fall round compared to the (first) summer round. To measure the interaction, the reviews and

replies were coded with a sequence number indicating the place in the conversation they belong to. The code was used to count the number of conversations by length. Table 5 shows that most initial contributions (84%) led to a conversation and that shorter conversations happened more frequently. The work in the PC could be done asynchronously, which limited the amount of interaction, but reducing the length of the review period reduces asynchronicity and may increase the online conversations. It would also be interesting to see whether conversations become longer when the PC becomes more familiar to the reviewers. Content analysis of the review contributions would be useful to differentiate between substantial (related to the evaluation criteria) contributions and more process-oriented contributions, which we may do in the next phase.

Table 5. Characteristics of the conversations.

Conversation Length *	Number of Conversations **	Total Contributions
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

* In number of contributions. ** Of these, some 30% are contributions from the AvH staff members. Source: Logfiles of the online platform.

The survey showed that about 80% of the respondents took note of the reviews done by others. About half of the interviewees said that they tended to stick to other PC members' comments, particularly to those of the 'first reviewer' who was assigned by the AvH staff. But they classified that as an advantage of the PC: taking notice of the opinions of others helps them form their own opinions and produce their own review and comments. An equally large group indicated that they made their own review and were not merely commenting on the reviews by others. Interviewees also mentioned another benefit of mutual influence: Reflecting on assessments of others can indeed lead to a change in their own assessment and help to reach agreement on who should be funded, which is seen as an important but not a necessary goal of the evaluation process. This shows that, unlike in the conventional review process, the PC reviewers are not working independently. Reading reviews from other members does influence PC members when writing their own reviews and comments, creating a risk of premature consensus that suppresses independent critical views on applications. Phenomena like groupthink are the mechanisms behind this risk (Janis, 1982), and interviewees indicated being aware of this possibility, but almost all asserted explicitly that it had not occurred. In this context, examples were given of diverging opinions. Being influenced does not necessarily lead to premature consensus.

Identifying the best applicants

What counts as 'the best applicants' depends on various criteria, of which one is the past scientific performance. Although reviewers generally claim to assess this by reading core publications of the applicants, the practice is different. Several PC members indicated in the interviews that publications received from the applicants were scanned rather than studied, and the AvH staff members also raised this issue in the interviews. The linguistic analysis of the review texts shows that discussions about publications focus on the length of

the publication lists, the impact of the journals, and the number of citations, and much less on the substance and findings. This applies to both the PC and the conventional process: The relative frequency of 'bibliometric' terms is in conventional reviews not different from the PC assessments (Table 4, row 'bibliometrics'). When observing the selection committees, we found the same: The discussions were frequently about the number of publications and citations and about the journal impact and hardly about the core publications the applicants had submitted.

We therefore use bibliometric data to answer the question of whether the PC is at least as useful for identifying the best applications as the conventional peer review mode. We do this for the two chemistry fields, where the use of bibliometric indicators is relatively undisputed. For each of the applicants (postdocs), we use the SciVal PP5 indicator[7] to measure performance: The PP5 indicator reports the share of papers in the oeuvre of a researcher that belong to the top 5% highest cited papers (by type of publication, field, and year of publication). The expected value is 5%, and on average the applicants score two to four times higher (Table 6). The average scores for the solid state chemistry applicants are somewhat higher than for the inorganic chemistry applicants, and this is in line with the committee scores the applicants received (Table 7).

Table 6. PP5 * score of the applicants, two chemistry fields **.

Field	2021		2022	
	Mean	Median	Mean	Median
Inorganic chemistry	14.2%	14%	11.3%	7.7%
Solid state chemistry	20.6%	11.8%	16.0%	10.0%

* PP5 = 5% highly cited papers. ** Peer Circle; N (2021): 28 (inorganic)–18 (solid state); N (2022): 17 (inorganic)–9 (solid state).

Table 7. Committee score of the applicants, two chemistry fields *.

Field	2021		2022	
	Mean	St. Dev.	Mean	St. Dev.
Inorganic chemistry	0.59	0.80	0.45	0.69
Solid state chemistry	0.84	0.91	0.63	0.95

* Peer Circle; N (2021): 28 (inorganic)–18 (solid state); N (2022): 17 (inorganic)–9 (solid state).

The next question is whether the selected applicants are on average better than the non-selected, which is a useful criterion to evaluate the selection process (Van den Besselaar & Leydesdorff, 2009; Bornmann et al., 2010). This is not always the case as we showed elsewhere (Van den Besselaar et al., 2023, p. 36), but one should remember that there are other main selection criteria such as the (contextualized) career of the applicant and her future perspectives which could be but are not necessarily within academia. The evaluation should therefore not focus only on the number of publications and their impact, and not unexpectedly, within the set of on average high performing applicants there are some with higher bibliometric scores that were not selected and some others with lower scores that did get a grant. In relation to this, the heterogeneity of applicants in terms of their home country may also play a role.

Perceived overall quality

The PC members were also interviewed and surveyed about how they perceived the overall quality of the PC reviews compared to the quality of the conventional reviews. Having been involved as a reviewer in different contexts, most PC members are able to compare the PC and the conventional process. We supplemented this self-reported approach with again an analysis of the review texts. It is important to note that the quality

of conventional reviews varies significantly and that 'the' quality of conventional reviews cannot easily be used as a standard for evaluating PC assessments.

A full review of a project proposal requires a complete understanding of all aspects of it, which most reviewers will not have, given the increasing complexity of research proposals. In the PC, one or two members address the evaluation dimensions as completely as possible, whereas the others add their partial assessment and their reactions to the contributions of others. The question is whether this may lead to incomplete and unbalanced reviews. One may argue that the PC cannot contain real technical specialists for each application that is reviewed. However, several interviewees emphasized that this is not a problem specific to the PC, since the conventional approach also fails to guarantee the reviewer's scientific proximity to the project. And as careers and future perspectives differ between countries, assessing those requires local knowledge which is more likely available within the larger PC than for individual reviewers.

The interviews show that almost all PC members found the reviews at least as comprehensive and detailed as the conventional reviews. While individual review comments may not always be as detailed as traditional reviews, the larger number of reviews and comments from multiple perspectives was seen as compensating for that. The more reviewers that have a look at an application, the more impartial, the more balanced and the better the overall evaluation. These opinions can be compared with the characteristics of the review texts just as we did above for the writing style. The depth and level of detail are expected to relate to the review's length, and Table 8 shows that the conventional reviews are longer.

Table 8. Length of the review texts and components.

Review Mode		Total *	CV	Core Publications	Project Proposal	Future Potential
Conventional	Average	1253	347 (28%)	292 (23%)	422 (34%)	192 (15%)
	CoV **		0.48	0.59	0.52	0.52
Peer Circle	Average	751	200 (27%)	141 (19%)	311 (41%)	99 (13%)
	CoV **		0.67	0.85	0.59	0.70

* Sum of numbers of the words used in the review. The profile text is not included, and the same holds for the summary texts, as these were not present in PC reviews. ** Coefficient of variance.

According to the interviewees, this was an effect of the more informal writing style we found in the linguistic analysis reported above. The PC members stated in the interviews that they focused less on the research project than they would in a conventional review. But in contrast to this, in the PC on average 41% of the review text is devoted to the project, whereas in the conventional reviews this is only 34%. An explanation for this unexpected finding could be that the PC reviews also include explanations of technical parts for those PC members who are less familiar with the subject of the project. But it also indicates that the PC is not neglecting the scientific-technical content of the proposed research. Furthermore, the PC allocates less text to the discussion of the core publications and the same share of the review to discuss the CV and future potential of the applicant (Table 8).

Summarizing, the interviewed PC members (as did the involved AvH staff) reported that the PC assessments have both depth and scope which is at least as good but probably better than conventional reviews. Indeed, a larger and more diverse group of reviewers is expected to produce less noise in their joint evaluation (Hong & Page, 2004). The results of the survey are consistent with the interview findings reported above. Some 90% of the PC members found the reviews to be of high quality, and 80% found them to be at least equally good as the conventional assessment (Van den Besselaar et al., 2023). Not all PC members answered these comparative questions, especially those who for the first time

were reviewing grant applications and therefore could not compare their PC experiences with earlier conventional reviewing.

Acceptance of the Peer Circle review procedure

Conventional peer review is seen by many as the cornerstone of the scientific enterprise and as its main governance process (Reinhart & Schendzielorz, 2024). Changing it may face resistance within the scientific community. But after the experiment, the majority (83%) of the PC reviewers were positive about the PC approach, preferring it to the conventional peer review procedure. Several arguments were brought forward to support this assessment. (i) Conventional reviews more often go into technical details of the project proposal only, instead of assessing the overall value of a grant proposal and of the applicant. (ii) Virtually all interviewees mentioned as an important characteristic the substantial number of peers contributing to the evaluation compared to the one or two in the conventional peer review. It makes the review process more transparent and objective, resulting in improved decision-making that is more easily accepted. As the number of evaluators increases, so does confidence in the outcome of the evaluation process. This is confirmed to an extent by (iii) some interviewed committee members who mentioned that having comments from a larger set of reviewers increases the confidence they have in the result of the review procedure. (iv) In the PC, the review outcomes are less influenced by the selection of individual reviewers. (v) The PC offers more and broader expertise. AvH funding programs are aimed at researchers from abroad, and assessing their quality, CV and career prospects requires knowledge of the country where the applicants come from. And among seven PC members, there is more often someone with the required local knowledge than among just one or two reviewers. (vi) The substantial size of the PC guarantees in most cases that the required scientific expertise is available to review the applications. (vii) The PC also functions as a self-correction mechanism. With around seven members, it is unlikely that none of the members would notice wrong assessments.

Overall, the PC members preferred the PC approach over the conventional mode of reviewing. The exception was some reviewers from the humanities, particularly in the PC on modern history, who preferred the conventional procedure. When asked in the interviews why they did, they focused very strongly on the need to assess the scientific content of the work of the applicants, and the idea of partial reviews did not appeal to them. The participants in the PC experiment are not a random sample from the scientific workforce in Germany. But the findings suggest that support for the new model may on average be strong, but that in some fields within the humanities this may be less univocal.

7. Conclusions and Discussion

The aim of the PC experiment was to test whether less labor-intensive peer review models are possible and whether they can lead to high-quality reviews. This evaluation of the first field experiment with the PC model was set up as a comparative case study, using a variety of data and triangulation of the results to come up with robust conclusions, despite the relatively small scale of the evaluation.

The case study leads to the following conclusions. Firstly, the PC as an alternative way of reviewing contributes to the alleviation of the reviewer scarcity. It requires substantially fewer reviewers, has a reasonable workload for the PC reviewers and saves time overall and likely also individually, a finding that is based on the interviews, the survey and the analysis of the logfiles. Although PC membership requires more time for reading, time is saved because of (i) the support PC reviewers get from reading the reviews and comments by other PC reviewers, and (ii) the more informal and authentic writing style that takes much less time than conventional review writing. And even if a PC member spends more time than someone writing a conventional review, for the scientific community it saves

time as many more conventional reviewers are needed than PC members for handling the same amount of grant applications. It is also more efficient in another way, as the problems of late and missing reviews are solved due to the way the PC is organized. The issue of missing reviews has been substantial as the required minimal number of two reviews for each application was far from realized.

Secondly, the PC members were satisfied with the quality of the PC reviews and mentioned various advantages, such as that there are more views on an application, that one is better able to differentiate between applications as one sees a set of them, and not only one as in the conventional review approach.

Thirdly, differences between the two modes of reviewing were found, among others a different emphasis on the evaluation criteria. A majority of the PC members communicated that the PC puts less emphasis on small technical details of the proposed project, which was seen as an advantage. The different writing style we found indicates a different relationship between the reviewers and the reviewed. The style of the PC reviews proves to be less formal and shows a more equal and authentic relationship between reviewer and reviewed. Which is something one should encourage between peers.

Finally, when asking the PC members for an overall assessment of the PC compared to conventional peer review, the large majority indicated that they prefer the PC. Acceptance of the PC may not become a problem. In this context, we also asked whether the PC members saw their work only as service to the community or whether they also got something from it. The latter was indeed the case. Respondents mentioned three benefits: (i) one gets an overview of what kind of topics are researched in their field; (ii) it becomes clear how applications are assessed; and (iii) together this helps in writing better applications oneself. These benefits are indeed based on PC members reviewing a series of applications and not just one.

Several improvements were suggested for the PC, such as including a grading and ranking of the applications within the Peer Circle, which would more clearly express the relative ranking by PC members of the applications. This may make the PC even more effective and efficient.

Discussion

Studies on grant peer review focus on reliability, bias and costs, and on alternatives like more structured reviews, remuneration of reviewers (Publons, 2019b; Tartari & Kolympiris, 2022), distributed peer review (UKRI, 2025; Brainard, 2025), anonymization (Witteman et al., 2019), hybrid lotteries (Feliciani et al., 2024; Horbach et al., 2022; Philipps, 2022), and more recently the use of AI in the review process (ERC, 2026). There are also studies with proposals to distribute research funding without peer review (Vaesen & Katzav, 2017; Bollen et al., 2014), but these have not been adopted or experimented with. This indicates the central role peer review still has in the scientific community. It also explains why mainly hybrid lotteries have been used, where the top and bottom applications are selected via peer review and a lottery is used for the gray area in between (Feliciani et al., 2024).

There are also hardly any experiments with remunerating grant peer reviewers (Tartari & Kolympiris, 2022), but many funders seem to pay reviewers who also function as panel or committee members (Publons, 2019b). Studies evaluating material incentives for grant peer review are scarce, and e.g., only 17% of surveyed grant reviewers mention remuneration as an issue (Publons, 2019b). Financial remuneration was not mentioned by PC members, who do report that they profit in other ways from their PC membership. They learn how applications are reviewed and learn how to write better applications. Apart from stressing the costs involved (Schweiger et al., 2024), other negative consequences were also reported, such as that offering material rewards to referees tends to decrease the quality and efficiency

of the reviewing process and that it may undermine moral motives that guide referees' behavior (Squazzoni et al., 2013).

Our study focuses on one of the most mentioned problems with peer review, that is invited reviewers rejecting the invitation due to a lack of time, leading to the above-mentioned possible solutions. An interesting one is the model of distributed review where applicants review a series of competing applications (Butters et al., 2025). The reviewer problem is solved, and one gets also enough reviews per application to alleviate the reliability problem, but it may bring new problems such as confidentiality and strategic grading. The PC does not have these disadvantages, and in the PC the number of reviewers involved for each application is also at the required level for reasonable reliability (Mayo et al., 2006; Marsh et al., 2008). We expect that if the PC members were to grade the applications, the inter-rater reliability should be fine. As in the PC every application is seen by many reviewers; the noise introduced in conventional reviewing by reviewer allocation is expected to become smaller. What also helps is that the PC has a somewhat more structured review using mandatory questions that could be much more refined, in line with relevant psychological research (Kahneman et al., 2021).

The PC seems a viable alternative review model, given the support it received during the experiment. As shown above, it contributes to alleviating the lack of reviewers while keeping peer review as the core mechanism. Furthermore, that most PC members agreed to continue for one or two more years indicates a positive reception. Based on the study reported here, the AvH has extended the PC experiment for a few more years and on a larger scale by implementing the Peer Circle for more research fields. In due time a larger evaluation study may improve our knowledge about alternatives for conventional peer review.

This study also has limitations that point to new research questions. The few critical PC members were from the humanities, and field differences in appreciation of the model should be further explored. Some of the analytical instruments can be improved, such as quality measurement for fields where bibliometric data are less accepted and measures for other evaluation criteria like quality of careers in their national context. By asking the reviewers to make explicit how they assess the 'soft' criteria, a more structured way of evaluating would become possible, and it would make explicit what should be covered in an application. Furthermore, the linguistic approach to extracting evaluation dimensions from review texts could be refined by taking field differences into account and by differentiating between substantial (related to the evaluation criteria) contributions and more process-oriented contributions. The use of other approaches like LLMs may be a fruitful further line of research. Gender bias was only briefly addressed in the study but this did not lead to clear findings due to a too-small N (Van den Besselaar et al., 2023). Finally, it would be necessary to test the PC also for other grant types (like thematic grants) besides the individual career grants we focused on in this study. Despite the need for further research, by collecting a wide set of qualitative and quantitative data and by using a variety of methods, we feel that the current findings are robust.

Author Contributions: Conceptualization, P.v.d.B.; methodology, P.v.d.B. and C.M.; validation, P.v.d.B.; formal analysis, P.v.d.B. and C.M.; investigation, P.v.d.B. and C.M.; resources, P.v.d.B.; data curation, P.v.d.B. and C.M.; writing—original draft preparation, P.v.d.B.; writing—review and editing, P.v.d.B. and C.M.; supervision, P.v.d.B.; project administration, P.v.d.B.; funding acquisition, P.v.d.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research project was financially supported by the Alexander von Humboldt Foundation (contract number 2021-400). Part of the research was done within the context of the GRANteD project, which has received funding from the European Union's Horizon 2020 research and innovation program (grant agreement No. 824574).

Data Availability Statement: Data for this study were confidentially provided by the AvH Foundation. They cannot be shared due to privacy, confidentiality, or ethical restrictions. Researchers interested in using the data should contact the AvH Foundation.

Acknowledgments: We thank the staff members of the AvH Foundation, and especially Michelle Herte for the cooperation and for providing information. Herte also provided a detailed description of the existing procedures in grant selection at the AvH Foundation. We also thank the Peer Circle reviewers and the committee members for taking the time to participate in the interviews and for completing the questionnaire. Comments by four reviewers on the previous versions helped to improve the paper. The authors used no AI-based tools.

Conflicts of Interest: The first author is affiliated with the Vrije Universiteit Amsterdam and with TMC Research. The second author is employed by TMC Research. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The Alexander von Humboldt Foundation enabled the research by providing all available data and by providing feedback on the design of the study and on the report (Van den Besselaar et al., 2023) about the study. However, all decisions about the design of the study, the data collection, the analyses and interpretation of data, text, and publishing the results were solely taken by the authors.

Abbreviations

The following abbreviations are used in this manuscript:

AvH	Alexander von Humboldt Stiftung (Foundation)
ERC	European Research Council
HFST	Humboldt Forschung Stipendium (Humboldt Research Fellowship)
LIWC	Linguistic Inquiry and Word Count tool
PC	Peer Circle
G-cases	Proposal to accept the application without further discussion in the selection committee
D-cases	Application that needs discussion in the selection committee
R-cases	Proposal to reject the application without further discussion in the selection committee

Notes

1 Public research funding consists of direct funding of universities and research institutes and competitive project funding. A main difference is that project funding is distributed based on an ex-ante assessment of individual research proposals, whereas direct research funding is often evaluated ex post, based on observed performance in the context of national evaluation systems (Zacharewicz et al., 2019). In both funding approaches, peer review plays a large role, despite the ample research results showing the problems of peer review. Competitive project funding is disputed in terms of costs and benefits (Schweiger et al., 2024; Dresler et al., 2023; Barnett, 2021) and in terms of what would be the optimal share of total research funding to be devoted competitive funding (Sandström & Van den Besselaar, 2018). Direct funding can also have—sometimes strong—competitive elements.

2 https://www.humboldt-foundation.de/en/, accessed on 12 May 2026.

3 HFST is the Humboldt Forschung Stipendium (Research Fellowship for postdoc and experienced researchers).

4 More details about the procedure in comparison to the existing one: Van den Besselaar et al. (2023).

5 The 2015 version was used (https://www.liwc.app/, accessed on 12 May 2026).

6 In practice, quite some conventional reviewers did send in their review too late or not at all, with the effect that only in 60% of the cases two reviews are available and in the other cases only one.

7 SciVal is bibliometric tool of the Scopus database.

References

Abramo, G., & D'Angelo, C. (2025). Peer review research assessment: Are the reviewers really experts? *Research Evaluation*, *34*, rvaf043. [CrossRef]

Aczel, B., Barwich, A. S., Diekman, A. B., Fishbach, A., Goldstone, A. L., Gomez, P., Gundersen, O. E., von Hippel, P. T., Holcombe, A. O., Lewandowsky, A., Nozari, N., Pestilli, F., & Ioannidis, J. P. A. (2025). The present and future of peer review: Ideas, interventions, and evidence. *Proceedings of the National Academy of Sciences USA*, *122*(5), e2401232121. [CrossRef]

Adam, D. (2019). Science funders gamble on grant lotteries. *Nature*, *575*(7784), 574–575. [CrossRef]

Adam, D. (2025). The peer-review crisis: How to fix an overloaded system? *Nature*, *644*, 24–27. [CrossRef]

Baccini, A., & Re, C. (2025). Is the panel fair? Evaluating panel compositions through network analysis. The case of research assessments in Italy. *Scientometrics*, *130*, 2093–2135. [CrossRef]

Barlösius, E., & Philipps, A. (2022). Random grant allocation from the researchers' perspective: Introducing the distinction into legitimate and illegitimate problems in Bourdieu's field theory. *Social Science Information*, *61*(1), 154–178. [CrossRef]

Barnett, A. (2021). *Funding schemes that cost as much as they reward*. Available online: https://medianwatch.netlify.app/post/hidden_funding_costs (accessed on 12 May 2026).

Belém de Oliveira Neto, J. (2024). The challenge of reviewers scarcity in academic journals: Payment as a viable solution. *Einstein*, *22*, eED1194. [CrossRef] [PubMed]

Bollen, J., Crandall, D., Junk, D., Ding, Y., & Börner, K. (2014). From funding agency to scientific agency; collective allocation of science funding as an alternative to peer review. *The EMBO Reports*, *15*, 131–133. [CrossRef] [PubMed]

Bornmann, L., Leydesdorff, L., & van den Besselaar, P. (2010). A meta-evaluation of scientific research proposals: Different ways of comparing rejected to awarded applications. *Journal of Informetrics*, *4*(3), 211–220. [CrossRef]

Brainard, J. (2025). Should grant applicants judge competitors' proposals? *Science*, *389*(6756), 120. [CrossRef]

Butters, A., Marshall, M. B., Pinfield, S., Stafford, T., Bondarenko, A., Neubauer, B., Nuske, R., Schwidlinski, P., & Denecke, H. (2025). *RoRi working paper no 17: Applicants as reviewers: Evaluating the risks, benefits, and potential of distributed peer review for grant funding allocations*. RoRi. [CrossRef]

Chubin, D. E., & Hackett, E. J. (1990). *Peerless science: Peer review and US science policy*. Suny Press.

Cicchetti, D. V. (1991). The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behavioral and Brain Sciences*, *14*(1), 119–135. [CrossRef]

Cole, J. R., & Cole, S. (1981). *Peer review in the National Science Foundation: Phase II*. National Academies Press.

Cole, S., Cole, J. R., & Simon, G. A. (1981). Chance and consensus in peer review. *Science*, *214*(4523), 881–886. [CrossRef] [PubMed]

Cotton, C. S., Alam, A., Tosta, S., Buchman, T. G., & Maslove, D. M. (2025). Effect of monetary incentives on peer review acceptance and completion: A quasi-randomized interventional trial. *Critical Care Medicine*, *53*(6), e1181–e1189. [CrossRef] [PubMed]

Dresler, M., Buddeberg, E., Endesfelder, U., Haaker, J., Hof, C., Kretschmer, R., Pflüger, D., & Schmidt, F. (2023). Effective or predatory funding? Evaluating the hidden costs of grant applications. *Immunology & Cell Biology*, *101*(2), 104–111. [CrossRef]

ERC. (2026). *The use of AI in grant proposal evaluation. Guidelines for ERC panel members and remote reviewers*. Available online: https://erc.europa.eu/system/files/2026-03/Use-AI-grant-proposal-evaluation.pdf (accessed on 12 May 2026).

Feliciani, T., Luo, J., & Shankar, K. (2024). Funding lotteries for research grant allocation: An extended taxonomy and evaluation of their fairness. *Research Evaluation*, *33*, rvae025. [CrossRef]

Gorelick, D. A., & Clark, A. (2025). Fast & Fair peer review: A bold experiment in scientific publishing. *Biology Open*, *14*(3), bio061982. [CrossRef]

Heger, C. (2025a). *Great expectations? Researchers in Germany are optimistic about the potential of grant lotteries* [Under review]. Department 2, Science Studies, DZHW (German Centre for Higher Education and Science Studies).

Heger, C. (2025b). *Ineffective, illegitimate, and inefficient? Normative arguments in the debate on grant peer review and grant lotteries* [Under review]. Department 2, Science Studies, DZHW (German Centre for Higher Education and Science Studies).

Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences USA*, *101*(46), 16385–16389. [CrossRef] [PubMed]

Horbach, S. P. J. M., Tijdink, J. K., & Bouter, L. M. (2022). Partial lottery can make grant allocation more fair, more efficient, and more diverse. *Science and Public Policy*, *49*(4), 580–582. [CrossRef]

Hug, S. E., & Aeschbach, M. (2020). Criteria for assessing grant applications: A systematic review. *Palgrave Communications*, *6*, 37. [CrossRef]

Janis, I. L. (1982). *Groupthink: Psychological studies of policy decisions and fiascoes*. Houghton Mifflin.

Jerrim, J., & de Vries, R. (2023). Are peer reviews of grant proposals reliable? An analysis of Economic and Social Research Council (ESRC) funding applications. *The Social Science Journal*, *60*(1), 91–109. [CrossRef]

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A flaw in human judgement*. HarperCollins.

Kaltenbrunner, W., Birch, K., & Amuchastegui, M. (2021). Editorial work and the peer review economy of STS journals. *Science, Technology, & Human Values*, *47*(4), 670–697. [CrossRef]

Markowitz, D. M. (2019). What words are worth: National Science Foundation grant abstracts indicate award funding. *Journal of Language and Social Psychology*, *38*(3), 264–282. [CrossRef]

Marsh, H. W., Jayasinghe, U. W., & Bond, N. W. (2008). Improving the peer-review process for grant applications: Reliability, validity, bias, and generalizability. *American Psychologist*, *63*(3), 160–168. [CrossRef] [PubMed]

Mayo, N., Brophy, J., Goldberg, M., Klein, M., Miller, S., Platt, R., & Ritchie, J. (2006). Peering at peer review revealed high degree of chance associated with funding of grant applications. *Journal of Clinical Epidemiology*, *59*(8), 842–848. [CrossRef]

Mocanu, M., Rusu, V. D., & Bibiri, A. D. (2024). Competing for research funding: Key elements impacting the evaluation of grant proposal. *Heliyon*, *10*(16), e36015. [CrossRef] [PubMed]

Mom, C., & Van den Besselaar, P. (2022). Do interests affect grant application success? The role of organizational proximity. *arXiv*. [CrossRef]

Oxford University Press. (2024). *Peer review survey: Results report*. Available online: https://static.primary.prod.gcms.the-infra.com/static/umbrella/document/Peer_review_survey_report_External_version.pdf?node=612e0870332719989117 (accessed on 12 May 2026).

Oxley, K. (2025). Meetings that matter: The dual benefits of panel peer review. *Research Evaluation*, *34*, rvaf047. [CrossRef]

Pennebaker, J. W. (2011). *The secret life of pronouns; what our words say about us*. Bloomsbury.

Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PLoS ONE*, *9*(12), e115844. [CrossRef]

Philipps, A. (2022). Research funding randomly allocated? A survey of scientists' views on peer review and lottery. *Science and Public Policy*, *49*(3), 365–377. [CrossRef]

Publons. (2019a). *Global state of peer review*. Available online: https://publons.com/static/Publons-Global-State-Of-Peer-Review-2018.pdf (accessed on 12 May 2026).

Publons. (2019b). *Grant review in focus*. Available online: https://publons.com/static/Grant-Review-in-Focus-web.pdf (accessed on 12 May 2026).

Reinhart, M., & Schendzielorz, C. (2024). Peer-review procedures as practice, decision, and governance—The road to theories of peer review. *Science and Public Policy*, *51*(3), 543–552. [CrossRef]

Roumbanis, L. (2019). Peer review or lottery? A critical analysis of two different forms of decision-making mechanisms for allocation of research grants. *Science, Technology, & Human Value*, *44*(6), 994–1019. [CrossRef]

Roumbanis, L. (2024). New arguments for a pure lottery in research funding: A sketch for a future science policy without time-consuming grant competitions. *Minerva*, *62*(2), 145–165. [CrossRef]

Sandström, U., & Van den Besselaar, P. (2018). Funding, evaluation, and the performance of national research systems. *Journal of Informetrics*, *12*(1), 365–384. [CrossRef]

Schweiger, G., Barnett, A. G., van den Besselaar, P., Bornmann, L., De Block, A., Ioannidis, J. P. A., Sandström, U., & Conix, S. (2024). The costs of competition in distributing scarce research funds. *Proceedings of the National Academy of Sciences USA*, *121*(50), e2407644121. [CrossRef] [PubMed]

Siemroth, C. (2024). Economics peer-review: Problems, recent developments, and reform proposals. *The American Economist*, *69*(2), 241–258. [CrossRef]

Simsek, M., de Vaan, M., & van de Rijt, A. (2024). Do grant proposal texts matter for funding decisions? A field experiment. *Scientometrics*, *129*, 2521–2532. [CrossRef]

Smith, D. S., Kennard, N. N., Du, T., & McFarland, D. A. (2025). How values and uncertainty shape scientific advance in peer review. *American Sociological Review*, *90*(5), 879–915. [CrossRef]

Snell, R. R. (2015). Menage a quoi? Optimal number of peer reviewers. *PLoS ONE*, *10*(4), e0120838. [CrossRef]

Squazzoni, F., Bravo, G., & Takács, K. (2013). Does incentive provision increase the quality of peer review? An experimental study. *Research Policy*, *42*(1), 287–294. [CrossRef]

Tartari, V., & Kolympiris, C. (2022). *Peer review for science funding: A review.* NBER white papers on research funding. Available online: https://www.nber.org/sites/default/files/2022-05/Peerreviewforsciencefunding.pdf (accessed on 12 May 2026).

Tauzik, Y. R., & Pennebaker, J. W. (2010). Psychological meaning of words. LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. [CrossRef]

Teplitskiy, M., Acuna, D., Elamrani-Raoult, A., Kördin, K., & Evans, J. (2018). The sociology of scientific validity: How professional networks shape judgement in peer review. *Research Policy*, *47*(9), 1825–1841. [CrossRef]

Tite, L., & Schroter, S. (2007). Why do peer reviewers decline to review? A survey. *Journal of Epidemiology & Community Health*, *61*(1), 9–12. [CrossRef]

UKRI. (2025). *Distributed peer review—Rules and guidelines*. Available online: https://www.ukri.org/wp-content/uploads/2025/02/ESRC-120225-Funding-Opp-UKRIMetascienceAIEarlyCareerFellowships-DPRRulesAndGuidelines.pdf (accessed on 12 May 2026).

Vaesen, K., & Katzav, J. (2017). How much would each researcher receive if competitive government research funding were distributed equally among researchers? *PLoS ONE*, *12*(9), e0183967. [CrossRef] [PubMed]

Van den Besselaar, P., & Leydesdorff, L. (2009). Past performance, peer review, and project selection: A case study in the social and behavioral sciences. *Research Evaluation*, *18*(4), 273–288. [CrossRef]

Van den Besselaar, P., & Mom, C. (2022). The effect of writing style on success in grant applications. *Journal of Informetrics*, *160*(2), 101257. [CrossRef]

Van den Besselaar, P., Mom, C., & Herte, M. (2023). *Evaluation of the 2022 Peer Circle experiment at the Alexander von Humboldt Foundation*. TMC Research.

Van den Besselaar, P., Sandström, U., & Schiffbaenker, H. (2018). Studying grant decision-making: A linguistic analysis of review reports. *Scientometrics*, *117*, 313–329. [CrossRef]

Witteman, H. O., Hendricks, M., Straus, S., & Tannenbaum, C. (2019). Are gender gaps due to evaluations of the applicant or the science? A natural experiment at a national funding agency. *The Lancet*, *393*(10171), 531–540. [CrossRef] [PubMed]

Yang, Z., Zhou, X., Jiang, Y., Zhang, X., Gao, Q., Lu, Y., & Yang, A. (2026). Human–AI complementarity in peer review: Empirical analysis of PeerJ data and design of an efficient collaborative review framework. *Publications*, *14*(1), 1. [CrossRef]

Zacharewicz, T., Lepori, B., Reale, E., & Jonkers, K. (2019). Performance-based research funding in EU member states—A comparative assessment. *Science and Public Policy*, *46*(1), 105–115. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.