Triangulation of Qualitative and Quantitative Methods in Panel Peer Review Research

Thamar Klein, Meike Olbrecht
Institute for Research Information and Quality Assurance, Germany klein

Abstract

More and more research grants worldwide are distributed based on decisions of review committees. But how do these peer review committees come to a decision and what affects decision-making processes when research programs are evaluated? Surprisingly little research has so far been devoted to this question. Most of the literature on peer review concentrates on the issue of whether editorial peer review lives up to expectations. Peer review of research applications has received much less attention and not much is known about expert panel evaluations.

This paper introduces the use of methodological triangulation within the ongoing project „Peer Review at the German Research Foundation. Researching into Expert Panels: the example of the Collaborative Research Centers“ which aims to contribute to the research on committee peer review and to provide information about strengths and weaknesses of the CRC-peer review process.

1. Introduction

Up until the 1960s most of the research money awarded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) went to individual grant programmes. Today, however, more than half of the budget is dedicated to coordinated programmes such as the Collaborative Research Centres (CRC). These are institutions established at universities for a period of up to 12 years that allow researchers to pursue an outstanding research programme, crossing the boundaries of disciplines, institutes, departments and faculties. “They facilitate scientifically ambitious, complex, long-term research by concentrating and coordinating the resources available at a university” (DFG [6]).

In contrast to individual grants programmes, which are evaluated by several individual reviews, excellence in CRCs is established through panel peer review. In panel peer review, a group of reviewers discusses a grant application and reaches a joint decision. This kind of approach to peer review gains internationally more and more importance. It differs from the traditional mail peer review inter alia by uniform rating obtained through an open communication process as well as by the lack of anonymity. Thus in panel peer review other and different effects impact on decision and judgment making than in mail peer review.

2. Research Rationale

Even though more and more research grants worldwide are distributed based on decisions of review committees, surprisingly little research has been devoted to this kind of peer review. On the one hand group decision making processes have been studied so far mainly in the experimental settings of game theory (Grimm and Mengel [9]; Luhan, Kocher and Sutter [15]) and social psychology (Kerr and Tindale [11], Linneweber [14]). On the other hand peer review research focuses on peer review processes of journals (see overviews De Vries et al [5], Weller [22]) and – much less – on the judgments of individual reviewers and/or the final decisions of funding agencies. Peer review of research applications has received much less attention (Langfeldt [13]) and not much is known about expert panel evaluations (Olbrecht, Bornmann [18], Lamont [12]; Obrecht, Tibelius and D’Aloisio [19]).

The project presented here deals with how the group review process affects the granting of CRCs. In all decision-making processes based on peer review, professionally competent, neutral reviewing that is capable of rendering critical assessment is the essence of sound, forward-looking decisions. One of the research objectives is to analyze the impact of group dynamics on individual assessment. The factors affecting the total judgment and possibly distorting effects are analyzed. Also more general questions as How do groups come to a common judgement? and What are the pros and cons of the Peer Review procedure in groups? are addressed.

3. Methodology

Triangulation has come to assume a variety of meanings (compare Bryman [4], Fielding and
For the purpose of this paper we define it as a multimethod research in which quantitative and qualitative research methods are combined “to provide a more complete set of findings than could be arrived at through the administration of one of the methods alone” (Bryman, p. 3 [4]). Rather than using triangulation solely as a technique for validation we employ it in order to enrich the data and to ensure a comprehensive and deeper understanding of the matter.

In order to elucidate our methodological approach we will first need to explain the panel peer review process.

Researchers who plan to establish a CRC are requested to prepare a preliminary proposal of about 100 pages. The German Research Foundation regards the pre-proposal as a basis for discussion, not as a condensed version of the final full proposal. So, subsequent alterations, such as including further projects or removing previously intended projects, may be made. If they succeed at this stage, they may hand in a full proposal of about 500 pages. About 60% of the preliminary proposals succeed. After that, the next phase of the selection process starts. 80% pass the review sessions on newly proposed CRCs.

Whenever a CRC is granted, it is evaluated every four years by a panel of peer reviewers on a two-day site visit. Reviewers are asked to evaluate the results achieved in the preceding funding period as well as the future research programmes and to give recommendations on the budget. This review is the basis for the funding decisions of the Grants Committee. The committee consists of scientists appointed by the DFG Senate and of representatives of the federal and state ministries responsible for research and higher education. Two scientific members of the Grant Committee participate in every site visit review.

Thus the review process consists of four parts: First the session on the preliminary proposal (which consists of about 100 pages), second a session of the Senate Committee on CRC, then the session on the newly proposed or continuing CRC (this proposal consists of about 500 pages) and finally the Grants Committee on CRC.

The review panel that takes part in the two-day evaluation of a newly proposed or continuing CRC consists of about 10 peer reviewers, 2 representatives of the Grants Committee and 2 members of the German Research Foundation. Each of the 10 reviewers is appointed to evaluate at least two of the CRC’s projects very closely.

The Two-Day site visit starts with an opening plenary session where each project head gives a 5 minutes presentation on the planned project. CRCs consist usually of about 15-20 projects. After each presentation the first short questions may be asked by the reviewers. The presentation of the planned CRC usually takes up all of the morning. After lunch the reviewers will have time to talk to the scientists who head the projects they were designated to review. This is called the poster session as most CRCs prepare posters for this event.

At 4 p.m. there will be the first closed discussion of the review panel members and the DFG representatives. This is where the actual discussion and marking will take place. Usually it lasts until about 9 p.m.

The morning of the second day is used for a discussion with the university administration, reviewers, and the participating scientists of the CRC. Here the very last questions that might have occurred during the discussion of the previous evening can be clarified. The morning ends with a second closed discussion of the review panel members and DFG representatives where the final marking is discussed. This marking will be communicated to the head of the CRC around lunch time. This is the end of the second phase.

Unfortunately the applicants have still to wait a few months – because the awards may only be granted by the Grant Committee which meets twice a year. Even if the review panel has marked the CRC as excellent, the Grant Committee may not grant the award due to lack of money.
data was analyzed and the research instruments were adjusted. In the following – and still ongoing – phase IV we will observe four review sessions on preliminary proposals and six sessions on full proposals for starting or continuing CRCs. Five of these CRCs come from the social sciences and humanities, another five from the life sciences. This division was carried out in order to account for possibly existing divergent evaluating schemes in different scientific cultures – especially as the German Research Foundation provides only a single list of questions and criteria for review sessions. The division also reflects the sciences with the most CRCs (life sciences) and the ones with the least CRCs (social sciences/humanities). Phase V is dedicated to data analysis and writing.

Thus the study includes 10 grant reviews for the implementation or prolongation of CRCs – each review process lasting two days – and eight advisory meetings for CRC pre-proposals. In the following we will only go into the reviews of grant applications and not into those of pre-proposals.

The project has three methodological characteristics. First: sequential triangulation of qualitative and quantitative methods. Second: inclusion of all actor-perspectives (applicants, reviewers, representatives of the Grant Committee, university (vice-)chancellors, governmental and DFG representatives). This will allow us not to oversee blind spots of methods or actors. And third: non-participant observation of all three stages of review (review sessions on preliminary proposals, review sessions of full proposals and the bi-annually meetings of the Grant Committee).

In our project we triangulate methods as well as sources. As already pointed out we asked for the perspectives of all people involved in the evaluation process. At the end of phase IV we hope to have carried out about 150-200 semi structured qualitative interviews. These include expert interviews with up to 12 reviewers per grant review (approx. a total of 110), interviews with 20 applicants for CRCs (the applications are handed in by teams of researchers. Two researchers per team were interviewed.), two - three expert interviews per review process with the German Research Foundation’s staff involved (approx. a total of 20 as some attended several review processes), about 20 interviews with representatives of the Grants Committee, as well as guided interviews with university (vice-)chancellors and governmental representatives (the latter take part in the final selection process of the Grant Committee after the panel peer review). Through the triangulation of different viewpoints towards the same panel peer review process we are able to get a more precise picture of this process.

The object of the interviews is to learn about the criteria panelists use – and their ranking – to decide on the fundability of CRCs, their views about the outcomes of the peer review process, as well as their experiences with and views about different types of peer review.

Besides the 200 qualitative interviews a quantitative online questionnaire of the panel peer reviewers from German Research Foundation who were not involved in the 10 observed cases will be carried out after the analysis of the qualitative data. The rich qualitative data will allow us to develop further meaningful questions. The quantitative questionnaire will be built on the basis of the expressive and nuanced qualitative findings in order to give breadth and representativeness to the data. It will allow us to quantify findings. Here qualitative and quantitative fieldwork is carried out in a mixed methodology design using sequential triangulation.

An example of concurrent triangulation of qualitative and quantitative methods is the use of cognitive methods (such as pile sorts and free lists) which are employed during the qualitative interviews. Beside the 200 expert interviews, the online questionnaire, and data gained from cognitive methods, direct observation of the 10 two day long peer review processes and 3 meetings of the Grants Committee has been carried out. The fieldnotes taken during the panel meetings, the official protocols of the panel meetings and transcripts of the taped interviews are analysed with the help of MAXQDA software for qualitative data analysis. Further data sources are applications, review instruments (e.g. guidelines of the German Research Foundation) and protocols.

### 3.1. Non-participant observation

The observation technique applied in this study is an open (the observed know about the observation), non-participant (the observers don’t take actively part in the situation under scrutiny), genuine (observation of inartificial settings) third-party-observation (no self-observation of observers).

As the German Research Foundation had asked us not to use video or audio recording devices during panel meetings we decided on non-participant observation, even though this approach is sometimes criticized for changing the observed people’s behavior (Hawthorne effect). Feedback of the German Research Foundation’s staff who had participated in CRC panel meeting with and without our presence indicates however that our presence did not have significant impact on the review process. By observing multiple sessions (10 grant reviews and eight advisory meetings) the risk of observer impact was minimized even further.

The non-participant observation conducted during the study’s explorative phase served to evaluate our research questions. First and foremost this was not about gaining data but about the reviewing of
research questions’ relevance, the study’s feasibility and possibly necessary modifications to the research design (compare Hauser-Schäublin [10]).

As the observable events during phase IV were generally verbal interactions – and as we were not allowed to use video or tape recording – we conflated the minutes of the meetings and the observational data into a single record on the basis of Spradley’s “Descriptive Question Matrix” (Spradley [21]). These records on observations and discussions are a substantial aspect of the interpretation of the standardized marking questionnaires (see below). Through the combination of both methods it becomes apparent which kind of assessment criteria will be brought into the evaluation procedure at what point in time and by whom – and who refrains from verbalizing their prior assessment. It also allows for an analysis of differences in evaluation and discussion cultures between humanities and life sciences. The observation enables us to gain an understanding of the social processes taking place during the panel peer review. It permits us to study the interactions and (sub)cultural rules of scientific communities – pursuing inter alia the following questions: Which (unexpressed) rules guide the evaluation process? Are there observable patterns and structures? What kind of problems occur and what kind of strategies are applied to solve them? Which (informal) techniques of exercising control become visible?

One of the advantages of non-participant observation consists in the possibility to experience and collect facts which insiders regard as self-evident and as not worth mentioning in interviews. Conditions, situations and facts which remain unmentioned in interviews because interviewees are unaware of them or do not want to bring them up for other reasons are thus recordable, too.

3.2. Archival data analysis

Archival data analysis is a further method that we will triangulate with the ones mentioned above and below. First and foremost, we deal with primary – and not secondary – documents. Thus we are mainly analyzing documents prepared by individuals who took part in the relevant events (Mogalakwe [17]). The most important documents which we analyze in the context of the study are protocols corporately composed by the German Research Foundation’s personnel and the two representatives of the Grant Committee who observed the evaluation process. These are of great importance as the granting decisions made in the Grant Committee are based on two things: First the oral presentation given by two representatives of the Grant Committee who observed the evaluation process and second on the protocols that have been made available for all members of the Grant Committee.

At the center of attention during archival data analysis stand those aspects of the panel peer review process that were considered note-worthy and that found their way into the protocols. Our objective is to find answers to the following questions: What kinds of aspects are introduced into the protocols? Which of these aspects are mentioned during the discussion of the Grant Committee? What kinds of aspects/criteria during the review process are disregarded and do not make it into the protocols?

In a further step these data are compared to observation protocols of the Grant Committee meeting. Are all of the facts mentioned in the reports discussed during the meeting? Is something left out? Are there aspects which are highlighted in specific ways? Which facts generate new discussions? Do the representatives mention aspects of the panel peer review process that were not put on record in the protocols?

The bottom line is to examine and to compare the criteria mentioned during the panel peer review process on the one hand with the criteria recorded in the official protocols and on the other hand with the criteria discussed at the Grant Committee meeting. In other words, we follow the approach of documentary analysis especially accentuated by McCulloch “to try to understand documents in relation to their milieu, or in other words to relate the text to its context.” (McCulloch [16])

3.3. Guided interviews and cognitive methods

As already mentioned above, about 150-200 guided expert interviews will be carried out with reviewers, applicants, representatives of the Grant Committee, staff from the German Research Foundation, university (vice-)chancellors and governmental representatives. The triangulation of different viewpoints of the same panel peer review process enables us to get a more precise picture of the different stakeholders’ perceptions.

Three subject areas are brought up during the interviews: 1) common knowledge of and experience with peer review, 2) common knowledge of and experience with the review system of Collaborative Research Centres, 3) questions with regard to the observed CRC. Additionally cognitive methods (such as freelistings and pile-sorts) will be applied. The latter serve to uncover group specific cognitive models, integrated frameworks and systems of concepts (also called cultural domains) with which the different stakeholders operate during the panel’s procedure. A cultural domain is a delimited space of knowledge provided with an inner structure which is shared by members of a particular “culture” (Borgatti [3], Schnegg and Lang [20]). This culture may be a panel culture or faculty culture for instance. Cognitive methods give answers to questions regarding both the sharing and distribution of
knowledge. The acquisition and analysis of this culture specific knowledge is based on the experiences of Cultural Domain Analysis (Alexiades and Sheldon [1]; Borgatti [3]; Bernard [2]). In our study we make use of freelistings, pile sorts, ranking and true/false statements as quantitative methods. The resulting data will be analysed using SPSS, Anthropac and VAP_Pilesorts.

3.4. Standardized marking questionnaire

In order to triangulate different points in time during the review process we developed additionally a standardized marking questionnaire. We asked the reviewers to evaluate the projects of a specific CRC on a scale from 1 (excellent) to 6 (not fundable) at different points in time during the peer review process. We measured and compared the markings of all reviewers at four different times. As our project examines decision making processes, we study the development of decisions during peer review from the first personal impression on the basis of the funding proposal, to the more developed assessment after the short presentations (at the opening plenary session) and poster sessions of the first day, over the consensual group decision, up to the final individual judgment after the panel peer review process.

Figure 2. Timeline

We asked the reviewers to provide a short appraisal of the CRC projects for the first time before the beginning of the two day long panel peer review process and for the second time after the poster session of the first day but before the first closed discussion of the review panel members at 4 p.m. During the closed discussion we took fieldnotes of the consensually given marks. At the end of the second day we received a completed questionnaire for the third time. This allowed us to see what impact the group discussion had upon the individual opinions.

4. Preliminary results

Our first – and not yet empirically sound – results show a trend toward group polarization. The term group polarization describes the tendency of groups to adopt a more extreme position subsequent to a group discussion than before a discussion. This is explained among other things through the incidence that people form more extreme positions in group situations than in individual situations as the initial opinions intensify through group discussion.

When comparing the data of the standardised marking questionnaire before group discussion with the marking given during the panel meeting, we were able to observe polarisation tending towards a more positive opinion during group discussion. One explanation for this trend towards a more positive marking was given during interviews. Some reviewers explained that only CRCs with overall excellent projects have a chance of being funded. So, whenever they really liked the idea of a CRC but felt that the CRC was endangered through a few projects that were only good, they felt inclined to shift them towards a better mark. Thus, to some extent a strategic calculation concerning the marking of the subprojects was mentioned. One reviewer notes:

“If the DFG [German Research Foundation] states that we only get excellent SFBs [CRCs] founded then you shift it to excellent. This may lead to the inflation of marks in some instances.”

4.1. Variance of individual evaluations

The analysis of the standardized marking questionnaire partly showed strong deviations between the individual opinions prior to the group discussion (t1, t2) and the group evaluation (t3) as well as a high variance within the individual judgments.

Figure 3. Group polarization

Figure 2 illustrates the evaluation of a subproject that was not recommended for funding by the peer review panel. In the following this particular subproject will be exemplified.

In figure 2 the single experts’ individual evaluations prior to (t1) and after the discussion (t3) are pictured via points. The marking as 1 represents an excellent project, whereas the marking as 6 represents a non-fundable project.

The mean value is symbolized through a square. The consensual group evaluation of the closed panel discussion (t2) is imaged as a square as well. At the
level of t1 there was a great diversity in marks: expert G1 rated the subproject with “excellent” (mark 1). By contrast, two other experts (G4 and G5) were of the opinion that it was just a “good” project (mark 5). Projects marked as “good” belong to the lowest marked projects within CRC evaluations. The only lower mark is “not fundable” (mark 6). The evaluation’s mean is the mark 3,6 – thus between “very good” and “very good to good”. After the discussion the evaluation picture is much more consistent and shows less variance between the individual judgments. Only the reviewer who marked the project as “excellent” prior to the discussion, holds on to the opinion (after the discussion) that this is a subproject worth granting even though it was evaluated as less good than before. Expert G3 ticked “not able to give an evaluation” after the discussion, which is why G3 is not shown at the level of individual evaluations after the discussion.

Due to the peer reviewers’ panel evaluation the subproject was not recommended as fundable. This decision was not predictable based on the mean value of the individual evaluations. What has influenced this change in mind of the reviewers? In order to answer this question, it is necessary to examine the discussion procedure closely. Here, the triangulation of methods proves to be of advantage, as the knowledge from the non-participant observation can complement the knowledge gained by the standardized marking questionnaire.

Surprisingly the observation protocol shows that there was little discussion about that subproject (all in all only four experts participated in it). However the two reviewers especially assigned to this project agreed about the project’s weakness. These two reviewers were identical with the two experts who marked the project initially as “5”. Therefore, the role of the reviewers who are especially assigned to subprojects will be examined in the following.

4.2. Group discussion and expertise

For every subproject there are at least two experts (representing experts) who are assigned to it and who invest extra effort into its evaluation. They are chosen by the German Research Foundation owing to their expertise within the subproject’s field of research. These experts function is to explain the subproject and their evaluation of it to the review panel.

The preliminary analysis of our guided interviews with reviewers reveals that, if the representing experts agree in their evaluation, the other group members tend to rely on those two experts judgment. Usually they agree to adapt or hold back their own evaluation if it deviates from the experts’ opinion because they assume that the representing experts are able to evaluate the particular subproject better than they are able to.

A discussion mostly occurs only if both presenting experts’ evaluations dissent from one another. If this is not the case, a group discussion generally does not take place. Langfeldt [13] arrives at a similar conclusion. She examined six different panels that dealt with evaluations of the Norwegian research landscape. The panels were differently composed in terms of their constitutions and organizational approach. Langfeldt [13] came to the conclusion that the decision-making process was characterized by a distinct task sharing and that there was little interaction between the experts. Solely in those cases where there was an overlapping of spheres of competence a discussion about evaluations happened. Langfeldt concludes from this that overlapping spheres of competence between experts seem to be a significant condition to further a discussion of scientific quality. Our interim results imply that overlapping competence is a necessary, but not sufficient condition for in depth discussions to happen. Possibly due to working efficacy, in depth panel discussions occurred mainly if there was dissent between the two appointed reviewers.

Nevertheless reviewer seem to prefer group peer review for the evaluation of CRCs, because they are in a position to jointly reevaluate their arguments, to weigh up the arguments against one another, and to make distinctions between important and less important arguments. An often mentioned drawback of German panel peer review is that it is a nonpaid time consuming procedure, which takes up much time and keeps reviewers from their own research. Many are however willing to put in the extra effort as they regard this kind of work on a give-and-take basis.

4. Conclusion

Triangulation provides us with a number of opportunities. First of all it helps to uncover unknown or unanticipated phenomena. Taking into account different view points is likely to produce new insights and data that do not fit the initial theoretical ideas on the subject. Thus it helps also to generate new research questions that may lead to a richer understanding of the research question. Different viewpoints, data sources and methods cross-stimulate each other and fertilize the data. Second it can also inspire the creation of new methods as for example our standardised marking questionnaire which is able to uncover and visualize the effect of the peer review process on the individual opinions of the reviewers. Last but not least it allows researchers to be more confident of their results.
5. Acknowledgements

First of all, we thank the German Research Foundation for the opportunity to observe all parts of the review process. We are especially grateful as not much research has been done on panel peer review due to issues of sensitivity and confidentiality. We gratefully thank all participants, reviewers, applicants, representatives of the Grant Committee, staff from the German Research Foundation, university (vice-)chancellors and governmental representatives who have contributed to our study. Last but not least we would like to thank our colleague Jeffrey Demaine for his comments on a draft of this paper.

6. References


