

Article

Ensuring Sustainable Evaluation: How to Improve Quality of Evaluating Grant Proposals?

Grażyna Wieczorkowska and Katarzyna Kowalczyk * 

Department of Managerial Psychology and Sociology, Faculty of Management, University of Warsaw, Szturmowa 1/3, 02-678 Warsaw, Poland; gwieczorkowska@uw.edu.pl

* Correspondence: kkowalczyk@wz.uw.edu.pl; Tel.: +48-660442293

Abstract: The principle of sustainable development is an obligation placed on all entities involved in the implementation and delivery of the structural funds made available not only by the European Commission but also by grant donors from all over the world. For this reason, when applying for a grant, proposals need to demonstrate the positive or neutral impact of the project on sustainable development. To be able to select projects that will ensure sustainability, we need to ensure the effective evaluation of the proposals. The process of their evaluation should be objective, unbiased and transparent. However, current processes have several limitations. The process by which grants are awarded and proposals evaluated has come under increasing scrutiny, with a particular focus on the selection of reviewers, fallibility of their assessments, the randomness of assessments and the low level of common agreement. In our studies, we demonstrated how some of those limitations may be overcome. Our topic of interest is the work of reviewers/experts who evaluate scientific grant proposals. We analyse data coming from two prominent scientific national grant foundations, which differ in terms of expert's selection procedure. We discuss the problems associated with both procedures (rating style of the reviewers, lack of calibration and serial position effect) and present potential solutions to prevent them. We conclude that, to increase the unbiasedness and fairness of the evaluation process, reviewers' work should be analysed. We also suggest that, within a certain panel, all grant proposals should be evaluated by the same set of reviewers, which would help to eliminate the distorting influence of the selection of a very severe or very lenient expert. Such effective assessment and moderation of the process would help ensure the quality and sustainability of evaluations.



Citation: Wieczorkowska, G.; Kowalczyk, K. Ensuring Sustainable Evaluation: How to Improve Quality of Evaluating Grant Proposals? *Sustainability* **2021**, *13*, 2842. <https://doi.org/10.3390/su13052842>

Academic Editor:

Katarzyna Liczmańska-Kopcewicz

Received: 1 February 2021

Accepted: 1 March 2021

Published: 5 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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 (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: institutional funding; science policy; sustainability; evaluation; grant proposals

1. Introduction

A vital part of academic work relies on evaluation. One of the most important elements of evaluation is peer review, acting as both a filter for selection and a quality control mechanism [1]. However, the system is not flawless, and the disadvantages of the peer review system have been widely discussed [2–7]. Assessing grant proposals is one area where peer review is often used; however, the empirical evidence on the effects of grant giving peer review is limited [8,9]. The process of grant proposal evaluation has also come under discussion, with particular criticism focused on the selection of reviewers and the fallibility of their assessments [10]; the occurrence of conflicts of interest between the reviewee and the reviewer [11]; randomness of assessments and low level of agreement among experts [3]; bias against innovative research [8,12]; theft of ideas [13]; and favouring a particular group of applicants [14]. These biases may distort the process of evaluation and influence the distribution on public funding, financing projects thus resulting in not achieving “development that meets the needs of the present without compromising the ability of future generations to meet their own need” [15] (p. 20). In other words, promoted proposals that underwent biased evaluation may not meet the criteria of sustainable

development, support the best science and contribute to funders' desired outcomes and the most effective achievement of sustainability.

Our topic of interest is the work of reviewers/experts who evaluate scientific grant proposals in governmental grant-making agencies. Their work is critical because, based on their evaluations, decisions are made on which proposals will receive funding, hence indirectly shape and affect the production of knowledge, as well as societal progress, economy, health and well-being.

Literature reviews present examples of many biases accompanying the review process [16–18]; unfortunately, the work of evaluators has rarely been analysed in a systematic way and little research has been done into the peer review of grant applications [19]. This is a problem, as such analysis could help to counteract some of the biases that have been identified.

In our paper, we analyse the work of the evaluators reviewing grant proposals for national bodies. Our datasets come from two prominent scientific national grant foundations, which differ in terms of expert's selection procedure. We focus on the first step of the evaluation—where grant proposals are being evaluated by a small group of reviewers (usually two or three) and then those with the higher scores are discussed at a full panel meeting. We present the problems associated with both systems as well as potential solutions to prevent them.

This paper is structured as follows. In the Theoretical Section, we discuss the evaluation process concentrating on its main elements: reference pattern, calibration, rating style and multiple objects evaluation. In the next Section 2, we provide empirical support for our theoretical considerations. First we compare two sets of reviews coming from two scientific national grant foundations. Then, in the experimental study, we test the bias called serial position effect, demonstrating that the order of evaluated grant influences its evaluation. In Section 3 we present a results and discussion and in Section 4 conclusions.

1.1. Evaluation Process

When considering a grant proposal, reviewers use an evaluation system provided by the contracting institution, which often takes a form of rating scales (e.g., determining the scientific experience of the applicant on a scale of 1–5 and evaluating of the innovative character of the project and the impact of its implementation on the development of scientific discipline on a scale of 1–3). Conducting such evaluation requires the reviewers to perform two tasks: (1) compare the evaluated grants with the quality internal standard; and (2) transform their own, private response scale, into the one provided by the contracting institution.

1.2. Activating Reference Pattern

The results of many tasks are easily measurable, e.g., number of items sold or the number of attracted customers. In such scenarios, a simple algorithm can be used to assess task completion. While for most material objects there are precisely defined standards (e.g., technical parameters), the process of objectivising the assessment of mental products (e.g., evaluating a grant proposal) is a major challenge. This activity requires reviewers to make a comparison with a certain reference pattern/standard which is tacit [20]. This reference pattern may have the form of: (1) a typical object (an averaged representation of all grant proposals previously reviewed); (2) an ideal object (the vision of an ideal grant proposal, not necessarily existing in reality); or (3) an exemplary object (e.g., a representation of a paper that was recently reviewed by the given expert).

Exactly which pattern will be activated depends on the reviewer's habitual choices (e.g., some always focus on the ideal pattern) and changing situational factors (e.g., last written review) as well as the mental and physical status of the reviewer: mood [21] and cognitive and psychoenergetic resources [22]. It might happen, however, that reviewers do not have a ready reference standard, e.g., when they are reviewing research proposals for the first time.

1.3. Calibration

Reference pattern begins to emerge during subsequent evaluation, so the previous grant proposal becomes a reference for the next one. It can be assumed that, at some point in the series, the reference pattern will be crystallised/calibrated. This in turn may change the average value of the pattern. This process is called calibration [23] and can be described as creating similar patterns in the reviewer's mind. Experts learn how to use an available evaluation system (a rating scale, grading system or binary accept/reject decisions) to rate objects. The anchoring effect also plays a fundamental role for the calibration process. In the classic sense, in the anchoring effect, the number specified at the beginning plays the role of a reference point (anchor) for further evaluations [24–26].

An attempt to control a reviewer's pattern seems to be important to ensure similar conditions for all evaluations. At the beginning of a rating series, experts often cannot predict the evaluating range of the objects they will have to rate (e.g., how good or bad the next grant proposal is). For this reason, they avoid giving the highest and lowest grades. The process of calibration manifests itself in an initial avoidance of extreme categories [27,28]. This is particularly problematic for good or bad proposals, especially when they are evaluated in sequence and reviewers cannot change their assessment by going back to the proposal and re-evaluate it [29].

1.4. Rating Style

Another key factor in evaluation processes is the reviewer rating style. When evaluating grant proposals, the reviewer needs to assess the quality of a grant proposal often using numeric scales. Even though the scale is precise (e.g., 1–5), reviewers might transform their own, private rating scale into the one provided by the contractor. We can identify reviewers who use the whole range of scale (using all points, i.e., 1, 2, 3, 4 and 5), those who use only two values (e.g., 2 or 3) and others who avoid extreme endings. This is a manifestation of a person-specific rating style that can be expressed through: (1) tendency to lenient/severe assessments [20,30]; and (2) lacking of differentiation through dimensions, namely the halo effect [31–33]. Lenient evaluators are referred as zealots (they accept a lot) while severe ones as assassins (they reject a lot) [34]. The reasons for leniency/severity of the evaluators are the differences in the way they process information, their personality traits, individual features and situational conditions [16].

The influence of rating style on the assessment has been presented in many studies: in the work of examiners evaluating the work of medical candidates [16], in ratings received from police officers, nurses and social workers [35] and aircraft mechanics [36]. It should also be noted that evaluator rating styles can be assessed only when we have a sufficient number of evaluations provided by an individual.

1.5. Single Versus Multiple Evaluation

All grant proposals within a certain panel can be evaluated by the same set of reviewers or by a random set of reviewers (a different set of reviewers for each grant proposal). With a random set of reviewers, we cannot control what reference standard is activated in the mind of an expert, which may influence their evaluation.

Marsh, Jayasinghe and Bond [37] compared assessments of reviewers evaluating one or more proposals. They found that ratings of those who evaluated at least three projects were more cohesive with the evaluations of other reviewers of the same project and more correct (i.e., similar to the final grades). This suggests that conducting more than one evaluation helps a calibrated pattern to emerge. Interestingly, ratings of those who evaluated only one project were more severe on average than ratings of those who evaluated more proposals. In a group of evaluators who have evaluated ten or more projects, a more consistent rating style (thus, their tendency for severe/lenient evaluations) was observed.

Although previous studies have established that the assessment of grants is often subject to relatively low inter-rater agreement, the impact this variability has on actual

decisions about grant funding has received less attention [19]. To determine the inter-rater agreement, we need to have access to multiply evaluations of the same reviewers. A lack of information about the reviewers' rating styles may lead to wrong decisions. This is an example of adverse selection of reviewers [38].

Thus, we suggest that, to minimise the negative effects of rating stale and different comparison pattern, all projects within one panel should be evaluated by the same set of reviewers.

2. Methodology

2.1. Study 1: Comparison of Two Evaluation Systems

The aim of the first study was to compare two datasets of grant reviews coming from two national foundations:

- (1) The National Centre for Research and Development awards grants within the LEADER programme of up to half a million dollars per project (\$13 million in total).
- (2) The Foundation for Polish Science grants one-year scholarships to young researchers within the START 100 programme, amounting to around \$8500 per project.

The main difference in the evaluation system is the selection process employed by the reviewers. In the START project, the majority of reviewers (67%) evaluated one project only. In the LEADER project, the same three reviewers evaluated all projects in a certain panel.

2.1.1. START Programme

In the START programme, 126 proposals were competing for the scholarships. The evaluation procedure was structured as follows: First, a set of three reviewers were selected separately to evaluate each application on a 100-point scale. Then, another set of experts formed a committee that read the reviews and conducted an interview with the leaders of the projects. To be interviewed, the project needed to receive 65 points (an average across all three reviews). Overall, 252 reviewers were involved in the evaluation process of 126 projects (among whom, 67% evaluated one project only).

When we looked at the difference between the maximum and minimum ratings given to the same project, we found that for almost 40% of the projects the difference between expert evaluation exceeded 30 points on a 100-point scale. This discrepancy between the ratings is an indicator of the rating style and a lack of an agreed basis for reviews (see Figure 1).

Based on the 65 points criterion, only 64 applicants were invited to present their projects to the committee. If, for a scientific reason, we excluded the ratings of the most severe reviewer out of three, the ranking would change. When the mean was computed based on the rating of two reviewers, an additional 25 project leaders would be able to present their project to the committee, which would increase the number of invited presenters from 64 to 89.

We repeated the same analysis with changed threshold—increasing the criterion to 70 (see Figure 2). Excluding the most severe reviewer would result in 17 more invitations to present proposals. Of course, the strict reviewer could be right and two others mistaken in their evaluation, but otherwise fallible evaluation undermines confidence that projects deserve funding will receive it.

2.1.2. LEADER Programme

In the LEADER programme, all proposals were reviewed by the same set of reviewers for each panel. The number of evaluated projects depended on the field (mathematics, physics, economics, chemistry and biology) and ranged from 16 to 59. There were generally 18 experts.

Having ratings of many research proposals from each reviewer allowed us to determine their rating styles, which was indicated by mean, median and variability [16]. In Figure 3, we can see that the experts differed in terms of the mean ratings (some

were significantly more lenient than others) and variability (some differentiated more than others).

The presence of very lenient or very severe reviewers did not distort the assessments, as all competing proposals were evaluated by the same experts, i.e., severe reviewers influenced all the proposals in the same way.

When evaluating an object, an evaluator needs to compare it with another object which can take different forms (ideal proposal that exists in the reviewer’s mind, the last proposal that reviewer evaluated or the average of all previously evaluated proposals). After performing a couple of evaluations, a comparison pattern stabilises in the reviewer’s mind. The lack of such stabilisation poses a threat to evaluations, especially when reviewers conduct multiple reviews in a sequence. This risk is a cognitive bias called the serial position effect.

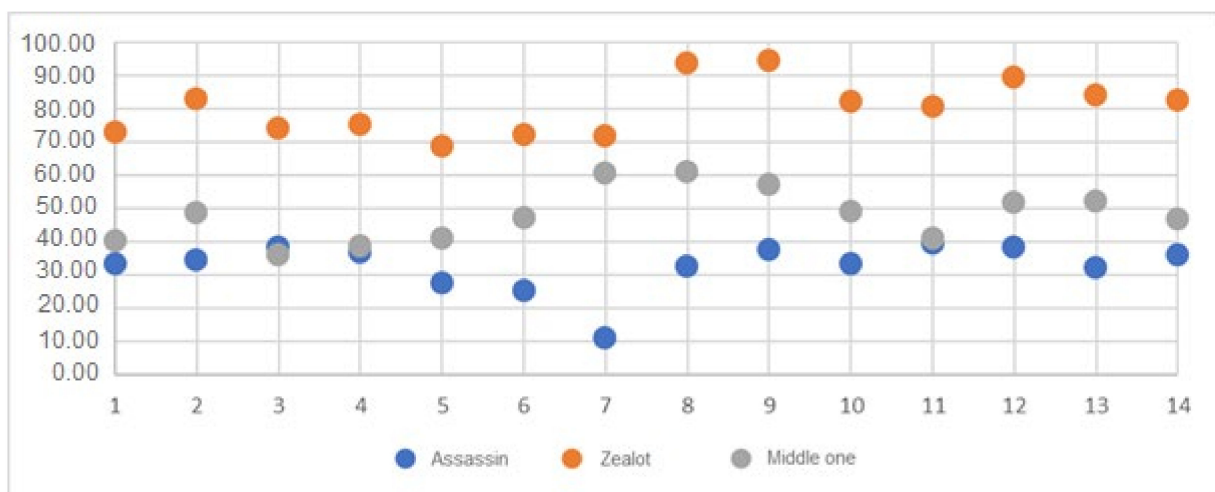


Figure 1. Differentiation of ratings (average, maximum and minimum value) of three reviewers for 14 projects that received at least 65 points.



Figure 2. Number of leaders getting a chance to present their projects with and without severe reviewers and the different selection criteria.

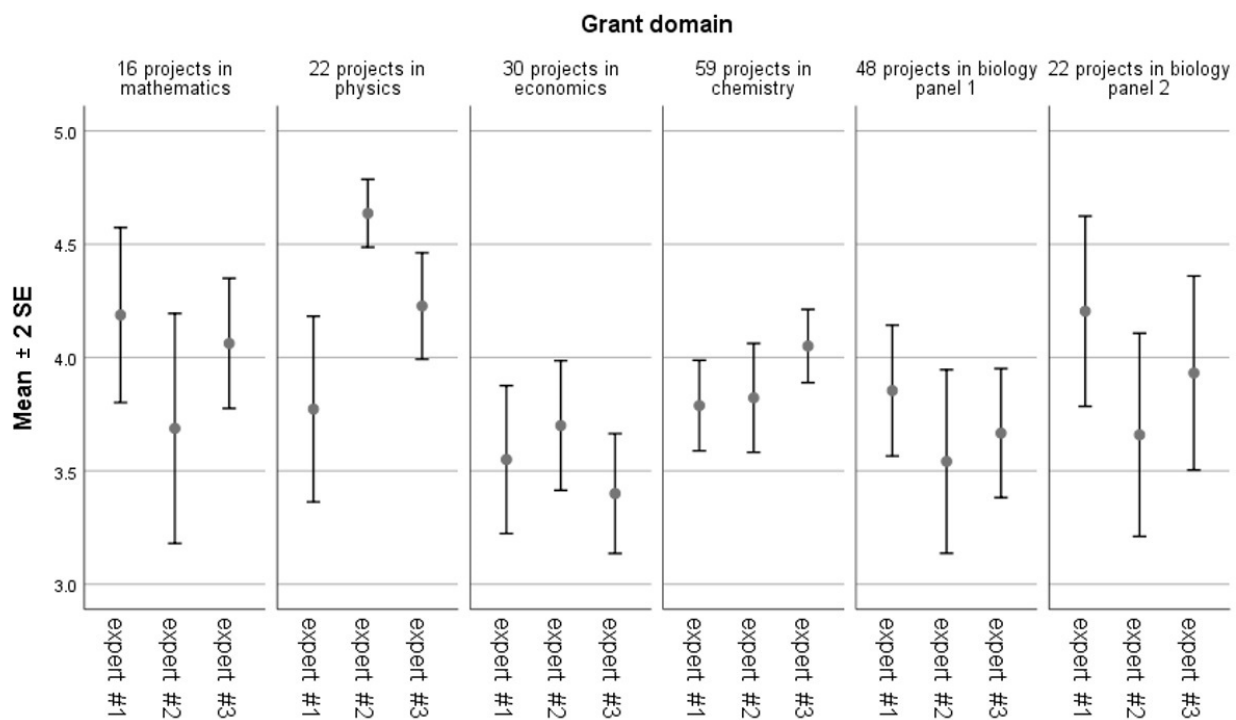


Figure 3. Comparison of rating style (mean and variability) of 18 experts (three for each of six domains) evaluating a series of grant proposals.

2.2. Study 2: Experimental Research on Serial Position Effect

When evaluating multiple objects, a comparison pattern emerges during the evaluation in the process of calibration. When calibrating, evaluators avoid extreme evaluations at the beginning of series. As a result, “good” objects are evaluated less favourably at the beginning of evaluation series, while “poor” objects are evaluated more favourably at the beginning of evaluations series as compared to the same objects evaluated at the end of a series. The serial position effect has been demonstrated in several studies [27,39–41]. It is worth noting that this effect occurs regardless of the degree of knowledge of the object under evaluation and the rating criteria [28].

In Study 2, we conducted an experiment demonstrating the occurrence of serial position effect in the context of conference abstracts evaluations. We also examined whether introducing a small break during the evaluation process would minimise the occurrence of this effect.

2.2.1. Participants

Eighty-six management students volunteered to participate in the study (70% women, aged 19–29, $M = 21.26$; $SD = 1.09$) in exchange for bonus points in one of their courses. The results of the experiment were later discussed during classes.

2.2.2. Procedure and Materials

An email with a link to an online survey was sent to students, inviting them to participate in the study. The participants took on the role of the evaluators of conference abstracts submitted to a conference on stress in the workplace, which was aimed at HR specialists and scientists. The abstracts concerned studies on the problems of occupational stress for different work groups (doctors, nurses, journalists, teachers, white-collar workers, etc.).

Each of the participants evaluated 15 conference abstracts on six dimensions on a seven-point scale: (1) overall abstract quality; (2) usability; (3) originality; (4) thematic compliance; (5) quality of presentation; and (6) general rating.

Based on the ratings collected in the preliminary study, the abstracts were classified into three categories: good, bad and average. A computer program randomly assigned the participants who filled out their consent form to take part in the study to one of the three experimental conditions. The scheme of the scientific procedure is in Table 1.

Table 1. The scheme of the experimental procedure.

E1	Good	Average	X	Poor
E2	Poor	Average	X	Good
E3	Poor	Average	BREAK TASK	Good

Group E3, during a break, was asked to evaluate six projects of conference logos on four dimensions on a seven-point scale: (1) originality; (2) thematic compliance; (3) precision of delivery; and (4) recommendation. After evaluating a whole series of abstracts, the participants indicated their age and gender.

2.2.3. Results

To test the impact of the order and the abstract quality on the evaluation process, a two-way ANOVA with repeated measurement on the last factor was performed, followed by the post-hoc Bonferroni post-test. The results are presented in Table 2.

Table 2. Effects of position and abstract quality on its evaluation. Dependent variable: Abstract evaluation.

SS		df	MS	F	<i>p</i>	Partial Eta Squared	Power
Within	15.24	55	0.28				
position (first, last)	12.96	1	12.96	20.42	0.001	0.271	0.993
Abstract quality (weak, good)	34.18	1	34.18	123.31	0.001	0.692	1.000
Abstract × order	2.86	1	2.86	10.32	0.02	0.158	0.884

All effects were statistically significant at the 0.05 significance level. The hypothesis predicting an interaction effect of the position and the quality of the abstract was supported. Poor abstracts gained and good ones lost when evaluated at the beginning of the series, while good abstracts gained and poor ones lost when evaluated at the end of the series.

As predicted, the number of extreme ratings was significantly lower ($t(87) = -4.35$; $p < 0.0001$) at the beginning of the series ($M = 2.75$; $SD = 2.83$) than at the end of the series ($M = 4.39$; $SD = 4.16$).

In the next step, the second hypothesis was examined, stating that the serial position effect could be minimised if a break were introduced, wherein the serial evaluation was interrupted to clean the operating memory. To assess the impact exerted by the break, the ratings given by Groups 2 and 3 for the best abstracts were compared. If break led to the reference pattern developed in the experts' minds vanishing, then the ratings of good abstracts in Group 3 would be lower than those in Group 2. Student's t-test for independent groups revealed no significant difference in the ratings of good abstracts conducted by Group E3 after a break with an additional task and by Group E2 working uninterruptedly without any break.

3. Results and Discussion

The aim our study was to underline the need for systematic analysis of experts evaluating grant proposals to increase the quality of their assessments. We compared evaluations of two datasets originating from the scientific grant competitions of two national agencies. The two datasets differed in terms of the reviewers' selection procedure. In the first case, all grant proposals were evaluated by different sets of reviewers, while, in the second case, all grant proposals from a certain domain were evaluated by the same set of reviewers. The first method does not allow controlling the comparison pattern which activates in the

reviewer's mind while reviewing. This might result in the lack of inter-rater reliability, understood as the extent of agreement between the independent reviewers [4]. As a consequence, this undermines confidence that projects deserving funding will receive it. This approach also does not allow for assessment of a reviewer's rating style, for example the tendency to severe/lenient assessments, which, as has been presented in many studies, influences the final evaluation [16,20,36]. Discordant evaluations among referees could make the review system unfair to authors whose manuscripts happened to be sent to an assassin or zealot reviewer [42] and thus should be controlled.

There are different methods for minimising the distorting impact of rating style, e.g., training aimed at increasing awareness of basic errors occurring during the review process. However, their usefulness is limited [43]. Giraudeau [44] proposed a simple method to identify discordant proposals, aiming to help track the proposals where evaluators disagree and further discussion is required. The first element of their method requires discarding any proposal with only one rating. One rating does not allow controlling the rating style of the evaluator or applying an algorithm that helps track evaluations that require further discussion.

To control the negative impact of an evaluator's rating style, we recommend a system where: (1) we have multiple evaluations from one evaluator; and (2) all grant proposals in a certain panel are evaluated by the same set of reviewers. If there are strict reviewers (assassin type), they will be strict for all of the evaluations in the process, and the same applies to those who are lenient (zealot type).

However, this approach also has limitations. If the evaluation is conducted in a series, reviewers are susceptible to the occurrence of the serial position effect. This distortion causes reviewers to avoid extreme ratings at the beginning of the series, which influences the evaluations of good proposals that are evaluated less favourably at the beginning of a series and bad proposals that are evaluated more favourably at the beginning of a series.

The serial position effect is explained by the calibration theory—evaluators learn how to use an available category system to evaluate objects [23]. In their first evaluations, evaluators avoid using very high or low grades as they do not want to violate the “internal consistency” of their evaluations. In other words, they are afraid to give maximum points to a proposal as they do not know the quality of the next one. After performing a couple of evaluations, a new comparison pattern is created in their mind, based on previous evaluations. We think that evaluators, prior to evaluating proposals, should receive a few proposals (i.e., the best and the worst from the previous competition), which would help to “calibrate” their mind and create a similar pattern to all. At the same time, it would give some insight into the reviewer's rating style and help to compute it. Similar solutions are used in the evaluation of applications in the American Graduate Research Fellowships programme of the National Science Foundation (GRFP). The calibration method raises some future research questions, for example: (1) How many proposals do reviewers need to evaluate to calibrate? (2) Is there a difference in reference patterns between zealots/assassins and more moderate reviewers?

Our attempt to minimise the serial position effect by introducing a small break task failed. Our intention was to clear evaluators' working memory and change the reference pattern in their mind. As a breaking task, we asked participants to evaluate six logo proposals. It might be that the task was too short, and the content of their working memory remained the same. In our future studies, we would like to examine if longer and different types of breaks would reduce the serial position effect. In our experiment, we used an online evaluation system that did not randomise the order of the evaluated abstracts and did not allow re-evaluating the proposals. It seems that allowing the two could help minimise the effect occurring when evaluating multiple proposals.

Finding ways to minimise the distortive effect of the described cognitive biases is important in the presence of an uncontrollable flood of information [45]. Over several decades, the number of scientific papers has climbed by 8–9% each year [46]. An explosion of scientific publications is accompanied by an increase in the number of cited references.

Overwhelmed by the volume of data to be parsed, the ability to accurately discriminate is increasingly important. Overflow leads to the shortening of time to concentrate on a single stimulus, succumbing to the influence of attention-grabbing stimuli and in turn adversely affecting information processing [47]. One of the consequences is heightened susceptibility to cognitive biases and reduced ability to defend against them. The tools and strategies that help to prevent the occurrence of biases are most wanted.

4. Conclusions

If we want ensure that the process of evaluation promotes proposals that are harmonising social, environmental and economical interest, reflecting values and ethics, the process of the evaluation itself need to be sustainable [48]. To achieve that, reviewers' work should be analysed. To prevent the distortive effect of evaluators' rating style and the selection of a lenient or severe evaluator, we suggest that: (1) evaluators should perform multiple evaluations, which would allow donors to control for their rating style and determine the zealots and assassins; (2) within a panel, all proposals should be evaluated by the same set of reviewers, which will allow controlling for the rating style as all of the proposals would be equally influenced by their rating styles; and (3) to create similar evaluation conditions, the evaluators should be calibrated prior to evaluation. To ensure the emergence of a similar comparison pattern, they should review the few best and weakest proposals from past competitions. We also suggest that within a certain panel all grant proposals should be evaluated by the same set of reviewers, which would help to eliminate the distorting influence of the selection of a very severe (who rejects a lot) or very lenient (who accept a lot) expert.

That effective assessment and moderation of the process would help ensure a more robust review of grant applications, providing for better use of resources and outcomes.

Author Contributions: Conceptualisation, G.W. and K.K.; methodology, G.W. and K.K.; software, K.K.; formal analyses, G.W. and K.K.; writing—original draft preparation, G.W. and K.K.; and writing—review and editing, G.W. and K.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by National Science Centre, Poland grant No. UMO-2016/21/N/HS4/00528: "Reviewing in Times of Overflow. Consequences for Managing Funds for Scientific Research".

Institutional Review Board Statement: Ethical review and approval were waived for this study by the Head of the Faculty of Management at the University of Warsaw due to low risk intervention on 4 November 2019.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data supporting the findings of this study are available from the corresponding author K.K on request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wager, E.; Godlee, F.; Jefferson, T. *How to Survive Peer Review*; BMJ Books: London, UK, 2002.
2. Jefferson, T.; Rudin, M.; Folsie, S.B.; Davidoff, F. Editorial Peer Review for Improving the Quality of Reports of Biomedical Studies. *Cochrane Database Syst. Rev.* **2006**. [[CrossRef](#)]
3. Bornmann, L.; Daniel, H.-D. The Effectiveness of the Peer Review Process: Inter-Referee Agreement and Predictive Validity of Manuscript Refereeing at *Angewandte Chemie*. *Angew. Chem. Int. Ed.* **2008**, *47*, 7173–7178. [[CrossRef](#)] [[PubMed](#)]
4. Bornmann, L.; Mutz, R.; Daniel, H.-D. A Reliability-Generalization Study of Journal Peer Reviews: A Multilevel Meta-Analysis of Inter-Rater Reliability and Its Determinants. *PLoS ONE* **2010**, *5*, e14331. [[CrossRef](#)]
5. Guthrie, S.; Guerin, B.; Wu, H.; Ismail, S.; Wooding, S. *Alternatives to Peer Review in Research Project Funding: 2013 Update*; RAND Corporation: Santa Monica, CA, USA, 2013.
6. Clair, J.A. Procedural Injustice in the System of Peer Review and Scientific Misconduct. *Acad. Manag. Learn. Educ.* **2015**, *14*, 159–172. [[CrossRef](#)]

7. Robson, K.; Pitt, L.; West, D.C. Navigating the Peer-Review Process: Reviewers' Suggestions for a Manuscript: Factors Considered before a Paper Is Accepted or Rejected for the Journal of Advertising Research. *J. Advert. Res.* **2015**, *55*, 9–17. [[CrossRef](#)]
8. Demicheli, V.; Di Pietrantonj, C. Peer Review for Improving the Quality of Grant Applications. *Cochrane Database Syst. Rev.* **2007**. [[CrossRef](#)] [[PubMed](#)]
9. Guthrie, S.; Ghiga, I.; Wooding, S. What Do We Know about Grant Peer Review in the Health Sciences? *F1000Research* **2018**, *6*, 1335. [[CrossRef](#)] [[PubMed](#)]
10. Tennant, J.P.; Ross-Hellauer, T. The Limitations to Our Understanding of Peer Review. *Res. Integr. Peer Rev.* **2020**, *5*, 6. [[CrossRef](#)]
11. Abdoul, H.; Perrey, C.; Tubach, F.; Amiel, P.; Durand-Zaleski, I.; Alberti, C. Non-Financial Conflicts of Interest in Academic Grant Evaluation: A Qualitative Study of Multiple Stakeholders in France. *PLoS ONE* **2012**, *7*, e35247. [[CrossRef](#)]
12. Boudreau, K.J.; Guinan, E.; Lakhani, K.R.; Riedl, C. The Novelty Paradox & Bias for Normal Science: Evidence from Randomized Medical Grant Proposal Evaluations. *SSRN Electron. J.* **2012**. [[CrossRef](#)]
13. Laine, C. Scientific Misconduct Hurts. *Ann. Intern. Med.* **2016**, *166*, 148–149. [[CrossRef](#)]
14. Murray, D.L.; Morris, D.; Lavoie, C.; Leavitt, P.R.; MacIsaac, H.; Masson, M.E.J.; Villard, M.-A. Bias in Research Grant Evaluation Has Dire Consequences for Small Universities. *PLoS ONE* **2016**, *11*, e0155876. [[CrossRef](#)] [[PubMed](#)]
15. Brundtland, G. *Our Common Future: Report of the World Commission on Environment and Development*; UN-Dokument A/42/427; UN: Geneva, Switzerland, 1987.
16. Dewberry, C.; Davies-Muir, A.; Newell, S. Impact and Causes of Rater Severity/Leniency in Appraisals without Postevaluation Communication Between Raters and Ratees. *Int. J. Sel. Assess.* **2013**, *21*, 286–293. [[CrossRef](#)]
17. Tamblyn, R.; Girard, N.; Qian, C.J.; Hanley, J. Assessment of Potential Bias in Research Grant Peer Review in Canada. *CMAJ Can. Med. Assoc. J.* **2018**, *190*, E489–E499. [[CrossRef](#)]
18. DeNisi, A.S.; Murphy, K.R. Performance Appraisal and Performance Management: 100 Years of Progress? *J. Appl. Psychol.* **2017**, *102*, 421–433. [[CrossRef](#)] [[PubMed](#)]
19. Graves, N.; Barnett, A.G.; Clarke, P. Funding Grant Proposals for Scientific Research: Retrospective Analysis of Scores by Members of Grant Review Panel. *BMJ* **2011**, *343*, d4797. [[CrossRef](#)]
20. Król, G.; Kowalczyk, K. Evaluation of Grant Proposals and Abstracts—The Influence of Individual Evaluation Style on Ratings. *Probl. Zarz.* **2014**, *12*, 137–155. [[CrossRef](#)]
21. Forgas, J.P.; Laham, S.M.; Vargas, P.T. Mood Effects on Eyewitness Memory: Affective Influences on Susceptibility to Misinformation. *J. Exp. Soc. Psychol.* **2005**, *41*, 574–588. [[CrossRef](#)]
22. Cole, S.; Balcetis, E. Sources of Resources: Bioenergetic and Psychoenergetic Resources Influence Distance Perception. *Soc. Cogn.* **2013**, *31*, 721–732. [[CrossRef](#)]
23. Memmert, D.; Unkelbach, C. Serial Position Effects in Evaluative Judgments. Current Directions in Psychology Science. *Curr. Dir. Psychol. Sci.* **2014**, *23*, 195–200. [[CrossRef](#)]
24. Galinsky, A.D.; Mussweiler, T. First Offers as Anchors: The Role of Perspective-Taking and Negotiator Focus. *J. Pers. Soc. Psychol.* **2001**, *81*, 657–669. [[CrossRef](#)]
25. Furnham, A.; Boo, H.C. A Literature Review of the Anchoring Effect. *J. Socio-Econ.* **2011**, *40*, 35–42. [[CrossRef](#)]
26. Kahneman, D. *Thinking, Fast and Slow*, 1st ed.; Farrar, Straus and Giroux: New York, NY, USA, 2013.
27. Unkelbach, C.; Memmert, D. Game Management, Context Effects, and Calibration: The Case of Yellow Cards in Soccer. *J. Sport Exerc. Psychol.* **2008**, *30*, 95–109. [[CrossRef](#)] [[PubMed](#)]
28. Fasold, F.; Memmert, D.; Unkelbach, C. Extreme Judgments Depend on the Expectation of Following Judgments: A Calibration Analysis. *Psychol. Sport Exerc.* **2012**, *13*, 197–200. [[CrossRef](#)]
29. Page, K.; Page, L. Alone against the Crowd: Individual Differences in Referees' Ability to Cope under Pressure. *J. Econ. Psychol.* **2010**, *31*, 192–199. [[CrossRef](#)]
30. Hoyt, W.T. Rater Bias in Psychological Research: When Is It a Problem and What Can We Do about It? *Psychol. Methods* **2000**, *5*, 64–86. [[CrossRef](#)] [[PubMed](#)]
31. Landy, F.J.; Vance, R.J.; Barnes-Farrell, J.L.; Steele, J.W. Statistical Control of Halo Error in Performance Ratings. *J. Appl. Psychol.* **1980**, *65*, 501–506. [[CrossRef](#)]
32. Cook, G.I.; Marsh, R.L.; Hicks, J.L. Halo and Devil Effects Demonstrate Valenced-Based Influences on Source-Monitoring Decisions. *Conscious. Cogn.* **2003**, *12*, 257–278. [[CrossRef](#)]
33. Dennis, I. Halo Effects in Grading Student Projects. *J. Appl. Psychol.* **2007**, *92*, 1169–1176. [[CrossRef](#)]
34. Siegelman, S.S. Assassins and Zealots: Variations in Peer Review. Special Report. *Radiology* **1991**, *178*, 637–642. [[CrossRef](#)] [[PubMed](#)]
35. Kane, J.S.; Bernardin, H.J.; Villanova, P.; Peyrefitte, J. Stability of Rater Leniency: Three Studies. *Acad. Manag. J.* **1995**, *38*, 1036–1051. [[CrossRef](#)]
36. Borman, W.C.; Hallam, G.L. Observation Accuracy for Assessors of Work-Sample Performance: Consistency across Task and Individual-Differences Correlates. *J. Appl. Psychol.* **1991**, *76*, 11–18. [[CrossRef](#)]
37. Marsh, H.W.; Jayasinghe, U.W.; Bond, N.W. Improving the Peer-Review Process for Grant Applications: Reliability, Validity, Bias, and Generalizability. *Am. Psychol.* **2008**, *63*, 160–168. [[CrossRef](#)] [[PubMed](#)]
38. García, J.A.; Rodríguez-Sánchez, R.; Fdez-Valdivia, J. Adverse Selection of Reviewers. *J. Assoc. Inf. Sci. Technol.* **2015**, *66*, 1252–1262. [[CrossRef](#)]

39. Bruine de Bruin, W. Save the Last Dance II: Unwanted Serial Position Effects in Figure Skating Judgments. *Acta Psychol. (Amst.)* **2006**, *123*, 299–311. [[CrossRef](#)] [[PubMed](#)]
40. Unkelbach, C.; Ostheimer, V.; Fasold, F.; Memmert, D. A Calibration Explanation of Serial Position Effects in Evaluative Judgments. *Organ. Behav. Hum. Decis. Process.* **2012**, *119*, 103–113. [[CrossRef](#)]
41. Antipov, E.A.; Pokryshevskaya, E.B. Order Effects in the Results of Song Contests: Evidence from the Eurovision and the New Wave. *Judgm. Decis. Mak.* **2017**, *12*, 415–419.
42. Chamorro-Padial, J.; Rodriguez-Sánchez, R.; Fdez-Valdivia, J.; Garcia, J.A. An Evolutionary Explanation of Assassins and Zealots in Peer Review. *Scientometrics* **2019**, *120*, 1373–1385. [[CrossRef](#)]
43. Schroter, S.; Black, N.; Evans, S.; Carpenter, J.; Godlee, F.; Smith, R. Effects of Training on Quality of Peer Review: Randomised Controlled Trial. *BMJ* **2004**, *328*, 673. [[CrossRef](#)] [[PubMed](#)]
44. Giraudeau, B.; Leyrat, C.; Le Gouge, A.; Léger, J.; Caille, A. Peer Review of Grant Applications: A Simple Method to Identify Proposals with Discordant Reviews. *PLoS ONE* **2011**, *6*, e27557. [[CrossRef](#)]
45. Czarniawska, B.; Löfgren, O. (Eds.) *Managing Overflow in Affluent Societies*, 1st ed.; Routledge: New York, NY, USA, 2012.
46. Landhuis, E. Scientific Literature: Information Overload. *Nature* **2016**, *535*, 457–458. [[CrossRef](#)] [[PubMed](#)]
47. Carr, N. *The Shallows: What the Internet Is Doing to Our Brains*; W.W. Norton: New York, NY, USA, 2010.
48. Silviu, A.J.G.; Schipper, R.P.J. Sustainability in Project Management: A Literature Review and Impact Analysis. *Soc. Bus.* **2014**, *4*, 63–96. [[CrossRef](#)]