Check for updates

OPINION ARTICLE

The jury is out: a new approach to awarding science prizes

[version 1; peer review: 1 approved, 1 approved with reservations]

Michael Hill

Swiss National Science Foundation, Berne, 3001, Switzerland

V1 First published: 03 Dec 2021, 10:1237
<https://doi.org/10.12688/f1000research.75098.1>
Latest published: 03 Dec 2021, 10:1237
<https://doi.org/10.12688/f1000research.75098.1>

Abstract

Research evaluation is often understood as something similar to a competition, where an evaluation panel's task is to award the most excellent researchers. This interpretation is challenging, in as far as excellence it is at best a multi-dimensional concept and at worst an ill-defined term because it assumes that there exists some ground truth as to who the very best researchers are and all that an evaluation panel needs to do is uncover this ground truth. Therefore, instead of focusing on competition, the Swiss National Science Foundation focused on active decision-making and sought inspiration in the deliberation proceedings of a jury trial for the design of a new evaluation procedure of an academic award. The new evaluation procedure is based upon fully anonymised documents consisting of three independent parts (achievements, impact and prominence). Before the actual evaluation meeting, the panel, which includes non-academic experts, pre-evaluates all nominations through a pseudo-randomly structured network, such that every nomination is reviewed by six members of the panel only. Evaluation decisions are based upon anonymous votes, structured discussions in the panel, ranking as opposed to rating of nominees and data-rich figures providing an overview of the positioning of the nominee along various dimensions and the ranking provided by the individual panel members. The proceedings are overseen by an academic chair, focusing on content, and a procedural chair, focusing on the process and compliance. Combined, these elements form a highly-structure deliberation procedure, consisting of individual steps, through which nominations proceed and which each either feed into the next step or into the final verdict. The proposed evaluation process has been successfully applied in the real world in the evaluation of the Swiss Science Prize Marcel Benoist, Switzerland's most prestigious academic award.

Keywords

Research Evaluation, Prize, Award, Review, Impact, Metrics

Open Peer Review

Approval Status

	1	2
version 1
03 Dec 2021	view	view

1. David Moher , University of Ottawa, Ottawa, Canada
2. Virginia Barbour , Queensland University of Technology (QUT), Brisbane, Australia

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the **Research on
Research, Policy & Culture** gateway.

Corresponding author: Michael Hill (michael.hill@snf.ch)

Author roles: Hill M: Conceptualization, Methodology, Project Administration, Supervision, Validation, Writing – Original Draft Preparation

Competing interests: MH is the Deputy Head of Strategy at the Swiss National Science Foundation. In this role he worked on the development of new evaluation procedures for the Marcel Benoist Foundation's Swiss Science Prize Marcel Benoist.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2021 Hill M. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Hill M. **The jury is out: a new approach to awarding science prizes [version 1; peer review: 1 approved, 1 approved with reservations]** F1000Research 2021, 10:1237 <https://doi.org/10.12688/f1000research.75098.1>

First published: 03 Dec 2021, 10:1237 <https://doi.org/10.12688/f1000research.75098.1>

Introduction

Prestigious academic prizes are usually awarded based on evaluation procedures where a group of experts ultimately decides who to award. However, during evaluation, experts may rely on inappropriate metrics,^{1,2} and discussion dynamics in the group may stray,^{3,4} which in turn can result in unfair processes^{5,6} and unsatisfactory outcomes.^{7–10} These circumstances are well known, and various initiatives, such as the San Francisco Declaration on Research Assessment (DORA),¹¹ the Leiden Manifesto,¹² the Metric Tide,¹³ the Hong Kong Principles¹⁴ and the h-group¹⁵ have highlighted them and provided guidelines on how to improve upon them. Inspired by these initiatives, the Swiss National Science Foundation (SNSF) has devised a novel evaluation procedure for the evaluation of academic awards. Essential to its innovative features is our interpretation of research evaluation less as a competition and more as an active decision-making process.

The task of an evaluation panel is often understood as that of finding the strongest applicant. However, while evaluation panels can usually discern the weakest and strongest contestants easily, they frequently struggle to delineate unequivocally, who from the midfield they might also want to champion or discuss and which one to ultimately single out as the winner.¹⁶ Comparison of scientific track records is multi-dimensional and the method of scoring individual dimensions as well as their respective weighting is controversial and often biased. Whereas in other competitions one can call upon millimetres and milliseconds as measures of merit, in research evaluation it is the formulation of a convincing argument within the evaluation panel, which ultimately calls the winner. Therefore, rather than thinking of research evaluation as something akin to a sports competition, it may be more useful to look to a different analogy, that of a jury trial.¹⁷

A jury does not have privileged access to an objective ground truth as a referee standing at the finish line might. Instead, the justification of a jury as a legal decision-making body is that it is comprised of peers and that, based on the trial proceedings, at this moment no other group of individuals are more informed and/or better qualified to make a decision upon the matter laid before them — even if a differently comprised jury might arrive at a different conclusion. The same is true for a panel of experts deciding which researcher to award. They cannot objectively establish who is the best scientist.¹⁸ Still, they can deliberate fairly and systematically upon why they might support one over another and can provide rational argumentation for this decision. It is important to note that a jury trial does not deny the presence of bias and predilection. Instead, it aims at controlling these forces by regulating deliberation and proceedings and making the final verdict democratic.¹⁹ Jury trials have their own set of challenges.²⁰ However, in the presence of biases and other human limitations they intentionally work toward a fair and rational outcome by systematically structuring, segmenting and transparently formalising their proceedings.^{17,19,20} This approach is not yet widespread among academic award committees, as a brief inquiry among the world's most established and prestigious academic prize institutions revealed (Box 1).

A proposal

Inspired by the analogy of a jury trial and the lack of such clear and transparent processes among academic prize committees, we propose a new award evaluation procedure. The procedure was also implemented in a first instance during the evaluation of the Swiss Science Prize Marcel Benoist (MBP), Switzerland's most prestigious academic award (see Application to the Swiss Science Prize Marcel Benoist,²¹). The new evaluation procedure is based upon the following three core principles:

Box 1: Evaluation guidelines for academic awards.

Upon inquiry, we found that for many prestigious academic awards there are either no detailed evaluation guidelines available or they are not publicly accessible. The following anonymised quotes are drawn from responses we received from some of the world's most prestigious academic award committees in response to the question:

"[...] May I ask you to possibly send me (a link to) a description of the proceedings, which take place during the actual jury meeting (i.e. is there first a pre-selection or consultation? How are all the nominations reduced down to the final winners? Are there rules or customs for this process?). [...]"

"The reason you do not find detailed information regarding the exact proceedings of the [...] Committee's work is that there are none."

"The committee is entirely free to structure their work the way they find appropriate"

"The Foundation keeps matters related to your query confidential"

"[...] according to the statutes of the [...] Foundation, the selection is treated confidentially"

"While the members of our prize Selection Committees are public, the proceedings are not"

"There are no written guidelines for how the committee decides upon a winner"

- Research evaluation is an active decision-making process by the evaluation panel. It is not the description of some objective ground truth by onlookers. The evaluation proceedings need to be structured to handle the complexity of the evaluation task appropriately.
- Evaluation and the documents under scrutiny should comprise of clearly delineated individual parts such that the verdict can be synthesised from the sum of many individual smaller steps. Assessment should not consist of unstructured, open-ended discussions that try to simultaneously consider all aspects of monolithic evaluation documents.
- Each step of the evaluation procedure must be transparent and well-defined, easy to understand with a clearly formulated aim and comprehensible outcome, which in turn should form the basis of the next step of the evaluation and/or feed directly into the final verdict.

In addition to these three principles, we also designed the process to rely on preparatory work such that the actual evaluation meeting itself can be streamlined and managed online. Flying in international experts for evaluation meetings is arguable neither necessary nor sustainable. The evaluation process is divided into three parts: nomination, pre-evaluation and finally the evaluation panel meeting itself. Each part is subdivided into individual steps.

Part 1: Nomination

To nominate a potential awardee, nominators fill in clearly structured nomination documents with the help of an interactive online platform. The interface provides all necessary definitions and information and helps the nominator organise the nomination correctly into three individual sections: “achievements *by* the nominee”; “prominence endowed *upon* the nominee”; and “impact originating *from* the nominee’s work”. Achievements are defined in line with DORA as the

“[...] actual work and output of the nominee. These may include important scientific publications, inventions, efforts, documented breakthroughs etc. Not regarded as achievements in this sense are prizes, awards and prestigious associations endowed upon the nominee (e.g. employment in famous universities, collaborations with famous people or publications in famous journals etc. [...]). Here, strictly only describe what the nominee themselves have actually done or produced.”

Prominence, in contrast, is defined as

“[...] prestigious distinctions and recognitions endowed upon the nominee by others as opposed to accomplishments attained by the nominee themselves [they] may include awards, titles, distinctions, nominations, and prestigious association such as employment in famous universities, collaborations with famous people or publications in famous journals etc (e.g., while the content of a seminal publication should be described in Achievements, the fact that it was published in a prestigious journal can be mentioned here, if you wish to do so).”

The prominence section is included because many nominators and panel members still today explicitly or implicitly rely on, independently google or outright demand such measures, especially when evaluating highly prestigious prizes. Rather than deceiving ourselves about these circumstances, we instead try to impose honesty and transparency in their use by strictly containing prominence information to one dedicated section. The definition of impact is the same as outlined in the assessment framework and guidance on submissions of the Research Excellence Framework 2014 (www.ref.ac.uk/2014), which defines impact as

“[...] an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia.”

Nominators are asked to keep their statements in the three sections concise and comprehensible to both an educated audience and non-expert members of the public, avoiding jargon. They are informed transparently about how the evaluation panel will use their statements. All information is provided directly within the online nomination mask such that no additional submission guidelines or further documentation are necessary. The same information texts are provided to the members of the evaluation panel. Separating the nomination text into three sections this way allowing the members of the evaluation panel to easily assess the actual achievements of the individual researcher independently of their prestigious associations, and the impact of their work without it being conflated with the work’s quality. Furthermore, it allows evaluators to easily refer to, evaluate and comment on these different aspects of a nomination independently and to compare them individually across nominations.

Any claims made by nominators within the achievements and impact sections have to be supported by references. No more than ten references, distributed freely across these two sections, can be used. Citations are entered into the text as bracketed numbers (e.g. [3]). The actual reference itself, however, only consists of information about the reference type (e.g. “journal article”, “book”, “radio interview”, “code” etc) and the respective abstract, synopsis or description, deliberately omitting any information about author, title, journal, publisher or publication date etc. Furthermore, upon submission of a nomination, the achievements and impact sections are anonymised such that they contain no names, gender or other identifying information but referred only to “the nominee” or the gender-neutral pronoun “they”. Any mention of, for example, institutions, collaborators or associations are also replaced with neutral references such as “at a Swiss university”, “with an established collaborator” or “as editor of an established journal” etc. Fully anonymised nominations can then be generated by separating the prominence section from the now anonymised achievements and impact sections. These fully anonymised nominations, consisting only of the achievements and impact sections with anonymised citations and without any mention of institutions or journal names etc, are then sent to evaluation panel members for pre-evaluation (Figure 1A).

Part 2: Pre-evaluation

Nominations are sent out to the individual panel members to review ahead of the panel meeting. The panel is comprised of at least seven and up to 11 members. In addition to international academic experts, it also includes two non-academic representatives of society and is age- and gender-balanced.

The nominations are distributed in a systematic, pseudo-random manner. They are assigned to panellists randomly albeit within the confines that every nomination is reviewed by six members of the panel only and that those six reviewers always include one of the two non-academic panel members and the panel’s topic-expert (e.g. all psychology nominations are read by the psychologist in the panel etc, Figure 1B). Furthermore, we ensure that each panel member’s individual collection of nominations to review is different to all others panel members’ sub-sets such that each nomination is compared to different nominations in every case and every nomination is compared to every other nomination at least once (Figure 1B). Importantly, this network distribution of nominations means that each panel member has to only pre-evaluate a sub-set of nominations (the number of which will depend on the total number of submitted nominations) instead of all of them, thus reducing the burden on the evaluators. In turn, the evaluation panel members have to commit to reading through their respective sub-set of nominations in full, including all the abstracts provided in the reference list, while also formally and explicitly committing to not seeking out additional information beyond what was provided (i.e. no googling of nominees etc).

The panel members are asked to pre-evaluate their respective sub-set of nominations in two steps. In the first step, they only receive the fully anonymised achievements and impact sections of the nominations in their respective sub-set

Figure 1. Pre-evaluation. (A) During pre-evaluation, each nomination (in this case, for example, “Nomination C”) consists of the fully anonymised achievement (Ach) and impact (Imp) sections only (without the prominence sections). References contain only a description of the reference type (e.g. “Journal Article”) and the respective abstract or synopsis. (B) “Nomination C” is then reviewed by a subset of evaluation panel members (here panellist 3 through 7), who each compare the nomination to a different set of other nominations respectively (in this case with some overlap). Each panellist ranks their set according to their personal overall preference (i.e. “who should win the prize?”, overall ranking (OR)) and additionally distribute gold (G), silver (S) and bronze (B) medals for the achievements and impact section respectively.

(i.e. without the prominence section, [Figure 1A](#)). These anonymised nominations have to be monotonically ranked according to the evaluation criteria and the individual panel member's personal assessment. Additionally, panel members are asked to assign an equal number of gold, silver and bronze medals across the individual achievements and impact sections respectively. If, for example, a panel member has to pre-evaluate 15 nominations in their sub-set, they would have to separately assign five bronze, five silver and five gold medals for achievements and for impact, respectively. The medals provide an intuitive way of generating tercile rankings of achievements and impact, in line with the overall ranking (i.e. first is better than second etc., [Figure 1B](#)). Finally, we ask panel members to guess who the nominated person is, if they believed they have an idea, to keep tabs on the extent to which anonymisation was successful.

All assessments throughout the whole evaluation procedure are always based on rankings instead of ratings. The advantage of using rankings (specifically weaker than/ stronger than) instead of ratings (generally weak/strong) is that they are internally normalised, thus mitigating the risk that a very generous or overly strict panel member might skew the evaluation. For example, by strictly specifying the number of gold, silver and bronze medals that can be allocated by a panel member, the chance of receiving more or fewer medals of a particular kind is independent of the individual panel members to whom the nomination is allocated. Instead, it is only a function of those individual panel members' evaluation of the respective nomination compared to the other nominations within their sub-set.

To further facilitate comparison across nominations, average and median values are calculated after pre-evaluation for each nomination based on the overall ranking and the medals it receives from the panel members who pre-evaluated it. The average overall ranking of a nomination is calculated as the mean value across its individual sub-set rankings. The average achievement and impact medals are calculated as the mean values across its individually assigned medals [defined as Gold = 1, Silver = 2, Bronze = 3 ([Figure 2A](#)), calculations of the median scores are done analogously]. These statistics are simple though calculating the mean across ranking data is, to some extent imperfect. However, presenting data in this very straight forward form allows all panel members to always easily and intuitively understand the statistics and their underlying data, which is crucial if they are to accountably argue based upon them ([Figure 2](#)).

In the second step of pre-evaluation, after returning their rankings, the panel members are provided with a detailed report and the full set of non-anonymised nominations. The analysis report provides three important results: (1) a detailed overview of all individual rankings and their mean and median positions across all nominations (see above); (2) a threshold calculated based upon these results, where only the nominations above this threshold are going to be considered further in the evaluation panel meeting; and (3) additional analyses of potential biases in the data.

The threshold is drawn based upon the following rule:

“Include only nominees with a mean AND median rank of seven or better in the overall ranking, OR a mean AND median rank of three or better in the achievement OR impact ranking.”

The panel members are then asked to indicate in writing whether they agree with the analysis and thresholding of nominees as described in the analysis report; or feel that one of the excluded nominees should still be included in the further evaluation. In the latter case, the panel member has to then also state in writing why the respective nominee should be singled out despite being below the analysis report's threshold. Any such argument has to also be presented personally at the beginning of the panel meeting so that the panel can vote to support or reject the proposition.

The additional analysis in the report consists of correlation tests between the panel members' rankings and additional parameters to illuminate any potential relationship between them, which might indicate a certain degree of bias. We compare the rankings of panel members from the same or similar research areas to those of panel members from other research areas than the respective nomination and we also correlate the panel member's rankings against the gender and age of the nominee. A correlation between rank and age would not necessarily be unexpected as more senior researchers have had more opportunity to excel. Nonetheless, these additional analyses can provide some insight into any potential systematic trends in the data, which might have to be addressed during the panel meeting, if they are of potential concern.

Along with the analysis report, the panel members are also provided with the full set of all original, non-anonymised nominations, now also including the prominence sections. The complete set of original nominations is offered to the panel members for their information only and they are not required to do any more work in preparation for the evaluation meeting at this point, except read the small set of nominations remaining above the threshold, in case they have not done so already as part of their own pre-evaluation sub-set. The panel members are, however, encouraged to read through all nominations and to revisit their own original sub-set, this time with added information such as name, gender, institution, and all the prestigious distinctions and recognitions outlined in the prominence section to the extent they feel necessary

Figure 2. Ranking data. (A) To provide an overview, the pre-evaluation data for the example of “Nomination C” from Figure 1 is listed in the table. Replacing the medals with numbers (gold = 1, silver = 2, bronze = 3), allows for the calculation of ranks (here only the average ranks are shown), which in turn allows for a positioning of each nomination relative to the others for the overall ranking (OR), the achievement ranking (ACH) and the impact ranking (IMP) respectively (plots). (B) In actual fact, the information provided in these rank plots needs to be substantially richer to be genuinely informative. It consists of the number of panellists who evaluated a nomination (n), the mean and median ranking value (red and blue dots respectively), the raw individual rankings from the panellists (gray dots) as well as the mean and median ranking (right y-axis) for each nomination (shown is the normalised overall ranking across 24 nominations based upon real-world evaluation plots but populated with random data for data-protection reasons, similar plots are also generated for the achievement and impact rankings respectively).

such that they can agree with the proposed thresholding and the resulting inclusion and exclusion of nominees. They can also use this information to argue against the thresholding if need be (see above).

Part 3: Evaluation meeting

The evaluation panel meeting is held online and led by two chairs. The academic chair is responsible for the content of the evaluation, and a separate, independent procedural chair is responsible for the evaluation process and compliance. The procedural chair guides the panel through the different steps of the evaluation. Whenever an individual nominee is discussed, all data on the respective nomination is displayed on screen: the panel members who pre-evaluated the nomination and the ranks and medals they assigned, the resulting position across all overall rankings and all achievement and impact medal rankings (Figure 2), the nomination synopsis, recusals, etc. The data allows for informed discussions (e.g. “nomination A ranked higher in achievements than nomination B”) and targeted questions (e.g. “why did panellist 7 give nomination C only a bronze medal for impact?”). Voting is conducted digitally and anonymously. All results are displayed as soon as the last vote is cast.

If a panel member at this stage strongly supports a nomination that they had originally ranked low, this becomes apparent quickly as all ranking information is displayed. Thus, if panellists change their mind because they now know who they are talking about, this can and should be discussed and challenged by the panel. The procedural chair is also tasked to look out for such discrepancies and highlight them, thus increasing transparency in the argumentation for or against a nomination.

All nominations, which are evaluated in the panel meeting (i.e. those that remained after thresholding), are discussed individually in detail to decide if they should be removed from the competition or forwarded to the final evaluation round. Each discussion starts off with a short plea against the respective nomination by the panel member who had pre-evaluated it and ranked it lowest compared to the other five pre-evaluators. Their statement is then followed by an argument for the nomination by the panel member who had ranked it highest during pre-evaluation. The two opening statements can then be commented upon by the remaining four pre-evaluators and then discussed further in the full panel. Finally, an anonymous vote is taken, whether the nomination should be put forward or whether it should be eliminated, before moving on to the next nomination. To keep the process efficient, only a limited number of nominations can be brought forward to the final discussion (e.g. three to five nominations).

While the initial discussions strictly focus on individual nominations only (i.e. should this nomination be put forward for this prize), the final session opens the floor to comparative discussions across all nominations (i.e. should this nomination be awarded over that nomination). It nonetheless is equally carefully structured to ensure fairness. All nominations and panel members are given equal air-time and the procedural chair intervenes when secondary or anecdotal information about a nominee is introduced. Such information is neither transparent nor fair and can work only for or against nominees familiar to members of the panel. The two non-academic panel members have the same rights and responsibilities as all the other panel members throughout the evaluation, in this step they do have one additional task, however, which is to confirm officially to the panel that in their view the remaining nominations are indeed impactful beyond academia, giving them a certain veto right in that regard. Once the comparative discussion converges on a few finalists, votes are cast to decide on the top two candidates and ultimately the winner. The first vote asks the panel members to individually rank all remaining nominations. The rankings reveal the final top two candidates from which a second vote between the two finalists declares the winner. The second vote is necessary to ensure that the majority of the panel ultimately cast their vote for the final winner.

Application to the Swiss Science Prize Marcel Benoist

The MBP is the most prestigious academic prize in Switzerland. Founded in 1920 it awards scholars who “[...] made the most useful discovery [...] that is of particular relevance to human life” and is therefore in its essence an impact prize. The award is rotated tri-annually through natural, biological and medical sciences, and the humanities and social sciences. Since 2018 the Swiss National Science Foundation is responsible for the evaluation of the MBP, in partnership with the Marcel Benoist Foundation.

The process presented here has been successfully applied to the evaluation of the MBP four times so far.²¹ The individual nomination sections consisted of: achievement, 800 words; impact, 800 words; and prominence, 500 words. Anonymisation took roughly four hours per nomination. However, not sending nominations out for external peer review also saved time, making the overall preparation efforts relatively efficient. Pre-evaluation and thresholding worked well and in one instance predicted the winner. Discussions during the evaluation meeting relied significantly on the information provided in the on-screen pre-evaluation data resulting in good discussion dynamics and transparent decision-making processes.

Compliance of individual panel members (e.g. no googling of nominations) and overall adherence to DORA guidelines (e.g. weighting the actual achievements of a nomination higher than the journal in which it was published) were some of the main challenges, which we encountered during the implementation of this process; they are, however, not unique or particular to this evaluation procedure. The process presented here heavily relies on preparation ahead of the actual evaluation meeting; due diligence in vetting nominations and good preparation especially by the procedural chair are crucial to ensure smooth and efficient evaluation. Overall, the experience with this evaluation procedure has been very positive and the SNSF will continue using it for the evaluation of the MBP and possibly other prizes in the future and some of its aspects are now also being adapted to other funding procedures.

Discussion

Many ideas, both tried and new, were combined in the design of this evaluation procedure^{22,23}: interpreting evaluation as more similar to a jury trial than a competition, segmenting and anonymising nominations, pre-evaluation, selectively assigning nominations to a network of panellists, using rankings instead of ratings, including non-academic experts in the evaluation panel, conducting voting anonymously and through dedicated infrastructure, displaying all nominee-specific and comparative evaluation data to the panel during discussions, dividing up the chairing duties between an academic and a procedural chair. Together, these innovations form a coherent, transparent and well-defined structure of individual steps, through which nominations proceed and within which panel members are supported in their deliberations.

Despite best efforts, for prestigious awards some nominees will almost always be familiar to at least some panellists while others may remain unrecognised. At the same time, maintaining the anonymity of nominations during panel discussions

is difficult and often not feasible. Acknowledging these challenges, we propose a pragmatic middle ground where nominations are kept anonymous during pre-evaluation but anonymity is then lifted in a controlled manner ahead of the meeting allowing for non-anonymous discussions.^{23,24} This approach provides the added benefit that it can also highlight biases originating solely from personal or prominence-based information as they stand in visible contrast to the pre-evaluation ranking based on the anonymised texts. Fully automated anonymisation for academic texts is, to our knowledge, not available or not reliable enough. However, a surge in privacy requirements, not least due to the General Data Protection Regulation (EU) 2016/679 have led to a proliferation of anonymisation tools,²⁵ which, for example in combination with [Research Organization Registry data](#), could, in future, be adapted for such purposes.

The evaluation process described here relies on nomination texts, which is unfortunate in those cases where a superficial nomination may disadvantage an otherwise very strong nominee. It is important to make nominators aware of this potential confound and the importance of a high-quality nomination on their part. Tampering with the process by, for example, adding additional information where needed, is not appropriate as it undermines the very essence of a nomination award. Such intervention leads down a slippery slope where one might as well just ask for the name of the person to be nominated and then leave the collection and selection of arguments to dedicated experts.

Nomination awards can also result in multiple nominations being submitted for some nominees but not for others. This can create an unfair advantage as multiple nomination texts provide opportunity to describe different accomplishments and to cite more works than just one text (although there will of course be some overlap between nominations). At the same time, multiple nominations may generate an implicit bias simply by suggesting that “more people think this person should win than that person”. Both situations need to be avoided, as the winner should be determined by the evaluation panel and not a majority vote from nominators.

To control for such circumstances, one can divide up multiple nominations for the same nominee among the pre-evaluators such that, for example, three panel members pre-evaluate only one and the other three the other nomination. During the second step of pre-evaluation, both nominations can then be provided along with all other non-anonymised information. As with the prominence section, panel members can now still change their mind based upon the additional information but any such changes are again highlighted due to their contrast to the original ranking data, allowing for the panel to call them out and discuss them transparently.

DORA encourages that evaluators read the work they are asked to assess as opposed to relying on short-cuts such as journal-derived metrics.¹¹ However, if evaluation organisers expect their evaluators to adhere to these guidelines, then they must also ensure that evaluators will actually be able to fulfil this mandate. You cannot commit to DORA and at the same time send out publication lists that are too long for evaluators to realistically read with due diligence. To address this problem, we only allowed for ten references to be cited across all three text sections (achievements, impact, prominence). However, for every ten nominations this would nevertheless add up to 100 papers to read, which is still a lot. Reducing the number of references even further may be difficult as academic awards, such as the MBP, are often given for a large body of work as opposed to a single project. Instead, we provided an abstract for each of the ten references and in turn asked the panel members to commit to reading all abstracts for every nomination in their sub-set. Providing abstracts instead of citations or full text publications also allowed us to maintain anonymity during pre-evaluation.

Researchers have a right to fair evaluation as they can make or break their careers.²⁶ Furthermore, small systematic biases can sum up to damage progress.²⁷ Despite best efforts, evaluators and evaluation processes are frequently — and to some extent inevitably — implicitly biased, and nominees may therefore not always be given their fair chance. The under-representation of women in academia generally and amongst academic awardees specifically is one example of the cumulative impact that biases at every stage of a scientific career can have.^{28–30} We believe many of the techniques discussed here could be transferred to other prize evaluations and assessments at other career stages, such as the hiring of faculty and the awarding of fellowships or grants. We hope that they may inspire change. High-quality and fair research evaluation requires not only a change in culture and a commitment to do better but most importantly also the actual implementation of fair and transparent processes.

Research evaluation functions as a gatekeeper of science. By assessing grant proposals in funding organisations, we determine who will be a scientist and what research will be conducted. By evaluating submissions to academic journals, we determine which research will be communicated to which audience. By evaluating nominees for academic prizes, we decide what research we want to celebrate and what academic role models we create for future generations. It is paramount to ensure that we can argue coherently and transparently how these important verdicts are reached in each of these situations.

Data availability

No data are associated with this article

Acknowledgements

The author wants to thank Michaela Strinzel for discussions, input on the manuscript and help with implementing these ideas in the evaluation of the MBP; Prof Matthias Egger for the opportunity, support, discussions, and input on the manuscript; Dr Katrin Milzow for support, discussions and feedback; Julius Mattern for help with [Figure 2](#).

References

- Müller R, de Rijcke S: **Thinking with indicators. Exploring the epistemic impacts of academic performance indicators in the life sciences.** *Res. Eval.* 2017; **26**: 157–168.
[Publisher Full Text](#)
- Smaldino PE, McElreath R: **The natural selection of bad science.** *R. Soc. Open Sci.* **3**: 160384.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lamont M: *How Professors Think: Inside the Curious World of Academic Judgment.* Cambridge, MA: Harvard University Press; 2009.
[Publisher Full Text](#)
- Olbrecht M, Bornmann L: **Panel peer review of grant applications: what do we know from research in social psychology on judgment and decision-making in groups?** *Res. Eval.* 2010; **19**(4): 293–304.
[Publisher Full Text](#)
- Merton RK: **The Matthew Effect in Science: The reward and communication systems of science are considered.** *Science.* 1968; **159**: 56–63.
[Publisher Full Text](#)
- Lincoln AE, Pincus S, Koster JB, et al.: **The Matilda Effect in science: Awards and prizes in the US, 1990s and 2000s.** *Soc. Stud. Sci.* Apr. 2012; **42**(2): 307–320.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Moher D, Naudet F, Cristea IA, et al.: **Assessing scientists for hiring, promotion, and tenure.** *PLoS Biol.* 2018; **16**: e2004089.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bol T, de Vaan M, van de Rijt A: **The Matthew Effect in Science Funding.** *SocArXiv.* 2018 April 23; **115**: 4887–4890.
[Publisher Full Text](#)
- Morgan W: **No Black Scientist Has Ever Won a Nobel- That's Bad for Science and Bad for Society.** *The Conversaton.* 2018. Accessed April 22, 2021.
[Reference Source](#)
- Lawrence PA: **Rank injustice.** *Nature.* 2002 Feb 21; **415**(6874): 835–836.
[PubMed Abstract](#) | [Publisher Full Text](#)
- DORA: San Francisco Declaration on Research Assessment: **Bladek | College & Research Libraries News.** May 2013. Accessed 27 Sep 2018.
[Reference Source](#)
- Hicks D, Wouters P, Waltman L, et al.: **Bibliometrics: The Leiden Manifesto for research metrics.** *Nature.* 2015; **520**: 429–431.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Wilsdon J, Allen L, Belfiore E, et al.: **The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management.** 2015.
[Publisher Full Text](#)
- Moher D, Bouter L, Kleinert S, et al.: **The Hong Kong Principles for Assessing Researchers: Fostering Research Integrity.** preprint. *Open Science Framework.* 2019.
[Publisher Full Text](#)
- Strinzel M, Brown J, Kaltenbrunner W, et al.: **Ten ways to improve academic CVs for fairer research assessment.** *Humanit. Soc. Sci. Commun.* 2021; **8**: 251.
[Publisher Full Text](#)
- Kaplan D, Lacetera N, Kaplan C: **Sample Size and Precision in NIH Peer Review.** *PLoS One.* 2008, July; **3**(7): e2761. Publisher: Public Library of Science.
- Hans VP: **Trial by Jury: Story of a Legal Transplant.** *Law Soc. Rev.* 2017; **51**: 471–499.
[Publisher Full Text](#)
- Fogelholm M, Leppinen S, Auvinen A, et al.: **Panel discussion does not improve reliability of peer review for medical research grant proposals.** *J. Clin. Epidemiol.* 2012 Jan; **65**(1): 47–52. Epub 2011 Aug 9.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Peter-Hagene LC, Salerno JM, Phalen H: **Jury decision making.** Brewer N, Douglass AB, editors. *Psychological science and the law.* The Guilford Press; 2019; (p. 338–366).
- Robertson CT, Shammas M: **The Jury Trial Reinvented: Boston Univ. School of Law, Public Law Research Paper No. 21-05.** March 1, 2021. SSRN.
[Reference Source](#) | [Publisher Full Text](#)
- Strinzel M, Egger M, Hill M: *Swiss Science Prize Marcel Benoist.* Hansson N, Angetter-Pfeiffer D, editors. 2021. in press.
[Publisher Full Text](#)
- Mayo NE, Brophy J, Goldberg MS, et al.: **Peering at peer review revealed high degree of chance associated with funding of grant applications.** *J. Clin. Epidemiol.* 2006, August; **59**(8): 842–848.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Sinkjaer T: **Fund ideas, not pedigree, to find fresh insight.** *Nature World View.* 2018.
[Publisher Full Text](#)
- Mosallanezhad A, Ghazaleh B, Huan L: **Deep Reinforcement Learning-based Text Anonymization against Private-Attribute Inference.** *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing.* 2019.
- Lauer M: *Anonymizing Peer Review for the NIH Director's Transformative Research Award Applications.* Extramural NEXUS, NIH; 2020.
- Casadevall A, Fang FC: **Is the Nobel Prize Good for Science?** *FASEB J.* 2013; **27** 12: 4682–4690.
[Publisher Full Text](#)
- Martell RF, Lane DM, Emrich C: **Male-female differences: a computer simulation.** 1996; 157.
- Gibney E: **What the Nobels are — and aren't — doing to encourage diversity.** *Nature.* 2018; **562**: 19.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Guthrie S, Rodriguez Rincon D, McInroy G, et al.: **Measuring bias, burden and conservatism in research funding processes.** *F1000Res.* 2019; **8**(851).
[Publisher Full Text](#)
- Guthrie S, Ghiga I, Wooding S: **What do we know about grant peer review in the health sciences? [version 2; peer review: 2 approved].** *F1000Res.* 2018; **6**: 1335.
[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 05 April 2022

<https://doi.org/10.5256/f1000research.78922.r126982>

© 2022 Barbour V. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Virginia Barbour

Office for Scholarly Communication, Queensland University of Technology (QUT), Brisbane, QLD, Australia

This is an interesting paper that describes the approach for the awarding of the Swiss Science Prize over the last four cycles of the Prize. The paper outlines the rationale for the approach, especially as it relates to more usual practices, and the challenges associated with the adoption and the ways that they have mitigated the challenges.

Comments:

1. The paper is valuable in itself as a description of an approach but in my view would benefit, if available, from an analysis of changes in nominations, and awardees of the prize in order to support the adoption of the new process.
2. The title gives a wrong impression of the paper. "The Jury is out" implies that something has not yet been decided. I appreciate there is a wish to have a catchy title, but it seems that a decision has been made to change this award process.
3. The paper does a good job of explaining the process that was adopted, and which is now in use, for the awarding of the Swiss Science Prize. The process of awarding science (and other academic) prizes is similar in many ways to the process of awarding research grants, though one notable extra step is the nomination process for prizes. The insights presented here may therefore be of interest to others involved in awarding prizes, as well as those involved in grants.
4. Although the introduction provides some background to the transition to the new process and the problems with current systems it would have been valuable to know whether any other model was considered.
5. In the discussion, the author touches on the importance of the nomination texts in determining who proceeds through for further evaluation. This seem such an important

step in the process that it would seem to be worth discussing further. It would be useful to know how many nominations are received and if that number has changed with the adoption of the new process.

6. Was there a specific rationale for change? For example, was there under-representation of women or other groups in the those who were nominated and also in who made it through to the final stages or the awardees? Related to this point, are there any data on whether there was any change in who was nominated, who was awarded a prize, or who ended up in the final deliberations? The paper would benefit from a presentation of this analysis. Is there any plan for future evaluation?
7. There is good recognition of how hard it is to fully anonymise applications and there is a good description of the steps that were taken. The author mentions that panel members are asked about whether they could guess who the nominees were. Are there any data on how successful anonymisation actually was?
8. It would be interesting to know more about the composition of the judging panel. Were there any changes to that composition between the previous and new models of judging?
9. Is there any plan to survey the panel members on their views on the new process?
10. The author mentions that there are plans to expand the process to other prizes. It would be useful to know if any changes in the process will be required for such an expansion.

Is the topic of the opinion article discussed accurately in the context of the current literature?

Yes

Are all factual statements correct and adequately supported by citations?

Yes

Are arguments sufficiently supported by evidence from the published literature?

Yes

Are the conclusions drawn balanced and justified on the basis of the presented arguments?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Publication ethics, journal peer review

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 11 March 2022

<https://doi.org/10.5256/f1000research.78922.r101950>

© 2022 Moher D. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

David Moher

School of Epidemiology and Public Health, University of Ottawa, Ottawa, Canada

Precis of paper

Hill describes a novel 'case' study for awarding Switzerland's most prestigious academic prize, namely, the Swiss Science Prize Marcel Benoist. The prize is rotated tri-annually through natural, biological and medical sciences, and the humanities and social sciences. The conceptual approach Hill describes is analogous to a jury trial in the legal system. The assessment for the prize has three core principles (and subdivisions): nomination; pre-evaluation; and an evaluation meeting (held online). The author indicates that this new process has been used four times for the prize.

My assessment

1. The topic is very important with insufficient investigation and thought, particularly on the part of funders. Too few funders are actively engaging in alternatives to the current unfortunate models. It is not a myth that after a grant application submission, colleagues email wishing me "the best of luck" or "fingers crossed". There is a recognition in the researcher community that how grants and/or prizes are awarded is a box of unclear processes; perhaps not reproducible.
2. This article very clearly describes the entire process of the prize evaluation. This is important and well done. There is a strong reliance on the work of DORA, advocating movement away from useless and non-evidence-based metrics and processes towards one that gets to the heart of research evaluations. Importantly the author discusses getting at the 'ground' truth. I wasn't sure what this term meant. Is it analogous to the 'gold standard'? I fear many readers won't understand the ground truth.
3. In part 2 - pre-evaluation, please provide some clarity about the time the nominations were sent to panel members in advance of the evaluation meeting. Similarly, please indicate whether other EDIA (equity, diversity, inclusiveness, and accessibility) beyond age and gender were considered.
4. A limitation of the paper for me is the lack of a discussion around the potential generalizability of using this approach for grant applications beyond prizes. Can this approach be considered for grant applications more generally? Did the author have an opportunity to survey the 'jury' as to their ease/difficulty using the approach, level of satisfaction with the process?
5. The pandemic has forced most of us to exist online. This has sometimes resulted in a variety of connectivity problems, postponement of meetings, etc. Can the author comment on connectivity problems and how different time zones posed any problems? I've often been asked to join meetings at 04:00. Not likely!

6. Another limitation of the paper was a lack of discussion around additional experimentation. For example, anonymizing certain parts of the process. In the 1980s, some medical journals were interested in whether double blinding manuscripts versus open was worth the effort in terms of quality of peer review – no was the general answer (e.g., Justice AC, Cho MK, Winker MA, Berlin JA, Rennie D. Does masking author identity improve peer review quality? A randomized controlled trial. *PEER Investigators. JAMA*. 1998 Jul 15;280(3):240-2. doi: 10.1001/jama.280.3.240. Erratum in: *JAMA* 1998 Sep 16;280(11):968. PMID: 9676668; van Rooyen S, Godlee F, Evans S, Smith R, Black N. Effect of blinding and unmasking on the quality of peer review: a randomized trial. *JAMA*. 1998 Jul 15;280(3):234-7. doi: 10.1001/jama.280.3.234. PMID: 9676666). Does the author think there is merit in thinking about experimenting with blinding versus open aspects of the described process for the prize evaluation?
7. The author points out various advantages of having the evaluation meeting online. Another one (not mentioned) is the reduction of a carbon footprint which perhaps meets an environmental/climate policy of a funder.
8. There is growing evidence of gender bias in the assessment of grant applications (e.g., Franco MC, Helal L, Cenci MS, Moher D. The impact of gender on researchers' assessment: A randomized controlled trial. *J Clin Epidemiol*. 2021 Oct;138:95-101. doi: 10.1016/j.jclinepi.2021.05.026. Epub 2021 Jun 9. PMID: 34118367; Witteman HO, Hendricks M, Straus S, Tannenbaum C. Are gender gaps due to evaluations of the applicant or the science? A natural experiment at a national funding agency. *Lancet* 2019;393:531–40. doi: 10.1016/S0140-6736(18)32611-4). I did not see a discussion as to whether the proposed approach might mitigate such bias(es).
9. Should the paper (end) with a clarion call for other funding organizations to initiate a discussion/pilot and evaluation of alternative ways to assess prizes and grant funding more generally? As indicated in the opening paragraph of the paper, several initiatives have called/alluded to this. Researchers deserve better than “best of luck”.

References

1. Justice AC, Cho MK, Winker MA, Berlin JA, et al.: Does masking author identity improve peer review quality? A randomized controlled trial. *PEER Investigators. JAMA*. 1998; **280** (3): 240-2 [PubMed Abstract](#) | [Publisher Full Text](#)
2. van Rooyen S, Godlee F, Evans S, Smith R, et al.: Effect of blinding and unmasking on the quality of peer review: a randomized trial. *JAMA*. 1998; **280** (3): 234-7 [PubMed Abstract](#) | [Publisher Full Text](#)
3. Franco MC, Helal L, Cenci MS, Moher D: The impact of gender on researchers' assessment: A randomized controlled trial. *J Clin Epidemiol*. **138**: 95-101 [PubMed Abstract](#) | [Publisher Full Text](#)
4. Witteman H, Hendricks M, Straus S, Tannenbaum C: Are gender gaps due to evaluations of the applicant or the science? A natural experiment at a national funding agency. *The Lancet*. 2019; **393** (10171): 531-540 [Publisher Full Text](#)

Is the topic of the opinion article discussed accurately in the context of the current literature?

Yes

Are all factual statements correct and adequately supported by citations?

Yes

Are arguments sufficiently supported by evidence from the published literature?

Yes

Are the conclusions drawn balanced and justified on the basis of the presented arguments?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Knowledge Synthesis; Open Science; Researcher Assessment

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research