Impact of Social Sciences Blog: Linguistic analysis reveals the hidden details of research grant proposal peer review reports

Page 1 of 2

Linguistic analysis reveals the hidden details of research grant proposal peer review reports

Despite peer review panels being the most common way of selecting applicants for research funding, little is known about how selections are made. New methods for large-scale text analysis allow for review panels' written reports to be analysed and studied for patterns. **Peter van den Besselaar** *and* **Ulf Sandström** *show how the frequency of positive and negative evaluation words correlate with applicants' final scores, and also reveal how panels' apparent conservatism means that the potentially groundbreaking, high-risk/high-gain research is unlikely to advance through the process and be selected for funding.*

Despite many studies showing the weaknesses of peer and panel review, it remains the dominant form of selecting applicants for positions and grants. However, quite what types of processes occur in review panels is not well known. Observation studies would be the way to investigate this, but these are very rare and those that are available are generally small-scale. It seems difficult for councils to allow for large-scale observations. One may argue that senior scientists have served on many panels and so should have enough insight; however, observing a panel is completely different from being on a panel while trying to do a good job.

Fortunately, with the availability of new methods for large-scale text analysis, an alternative way of studying panel processes is to analyse the written reports. In a study focusing on gender bias in grant allocation at the European Research Council, we obtained all review reports of both successful and rejected applicants (in total 3,030 applicants with between four and eight short reviews plus a short summary review report) for further analysis. We conducted a linguistic analysis of the reviews, using LIWC – a dedicated tool for such analysis. The tool distinguishes a variety of linguistic categories, each consisting of a large set of words. For our study we used linguistic categories assumed relevant for review texts, among others: negation words; negative evaluation words; positive evaluation words; superlatives, research project words; track record words. The question is whether we can predict the score an application gets from the relative frequencies of the linguistic categories in the reviews, and if so, what does this tell about the panel process?

Not unexpectedly, positive evaluation words and superlatives correlate positively with the score, and negation words and the negative evaluation words correlate negatively. Also, research project words and track record words correlate negatively with the score, so the more words used to discuss the research project and the track record of an applicant, the lower the score the applicant is given, on average. In fact, as literature suggests , more discussion about the applicant and application is a sign of disagreement and the higher the level of disagreement between panelists, the lower the score. Table 1 shows the regression results for the PI-score in the first round – the other analyses (PI in the second round and Project in first and second round) have similar results. As the table shows, the regression coefficient of the word categories that correlate negatively with the score are considerably larger compared to the word categories that correlate positively.

Model 5	Unstandardized		Standardized		Sig.	95.0% Conf Interv for B	
	B	Std. Error	Beta	t		Lower Bnd	Upper Bnd
(Constant)	3.423	0.062		55.151	0	3.301	3.545
Negation	-0.453	0.028	-0.472	-16.18	0	-0.508	-0.398
Negative Evaluation	-0.162	0.021	-0.189	-7.561	0	-0.204	-0.120
Superlatives	0.089	0.015	0.100	5.818	0	0.059	0.119
Proposal	-0.039	0.006	-0.092	-6.441	0	-0.051	-0.027
Track Record	-0.062	0.010	-0.088	-6.072	0	-0.082	-0.042
Certainty	0.258	0.044	0.081	5.841	0	0.171	0.345
Disagreement in panel	-0.190	0.039	-0.066	-4.923	0	-0.266	-0.115
Exclusion	0.053	0.022	0.064	2.426	0.015	0.010	0.096
Positive Evaluation	0.035	0.011	0.056	3.179	0.001	0.013	0.057
Positive Emotions	0.038	0.012	0.050	3.270	0.001	0.015	0.061

Adjusted R^2 = 0.463

Table 1: PI-score (first phase) by frequency of linguistic categories.

These results suggest that panels have a negative attitude: trying to find weak parts in applications to justify rejecting them, and not trying to find the breakthrough ideas that may be in an application, ideas that may bring science substantially further forwards and/or lead to radical societal impact. That panels do this is understandable; the number of applications – especially to the ERC starting grant scheme we analysed – is high, and the success rate is low. Panels have just a few minutes to discuss each application, and this decision-making pressure obviously leads to the use of efficient heuristics: an identifiable weakness means "out". This is a risky approach.

In fact, the starting grant is intended to fund groundbreaking, high-risk/high-gain research, which always has many uncertainties and aspects that are not yet fully developed. Identifying these projects requires an open-minded, risk-taking approach. Indeed, the panels seem to have adopted a conservative stance, looking more for certainty, rejecting proposals with some weak parts, and not going for the exciting, risky project which may fail but also has potential to make a real difference should it succeed. For example, one of the panels writes: "if the project would be successful, it would really change the field". The panel seems to hesitate over whether or not the field would take up the results of this project, which they call "a risk of the project". This is clearly a high risk; it's a high-gain proposal but the conclusion is that "the project may probably fail" and low scores are given, not even being high enough to advance to the second phase.

This rejection of worthwhile proposals is not a theoretical problem. In another study on the Netherlands Starting Grant scheme, we found that successful applicants did not perform better than an equally large sample of not-granted applicants with strong past performance. In fact, it was the other way around: after ten years the non-granted had performed better in most indicators we used. For example, on average they had a higher number of top-cited papers than the granted applicants. This means very qualified false negatives occur in the dominant panel based decision processes. What would be the policy lessons of this?

If this "negative attitude" is the effect of the enormous review pressure, then it may be an idea to organise funding differently. Why not distribute a large amount of the project money among all researchers based on a (age-related) performance indicator , without peer and panel review? A smaller amount may then be distributed following panel review with the intention of identifying and rewarding the really challenging ideas. For this latter procedure, past performance may help to select those who have shown themselves to be able to produce and develop real new ideas. Science is too important to institutionalise procedures which prefer normal problem-solving approaches above new ideas that may make the difference in the long term.

This blog post is based on the authors' co-written article, "Studying grant decision-making: a linguistic analysis of review reports", published in Scientometrics (DOI: 10.1007/s11192-018-2848-x), and on related work.

Featured image credit: Joyce McCown, via Unsplash (licensed under a CC0 1.0 license).

Note: This article gives the views of the authors, and not the position of the LSE Impact Blog, nor of the London School of Economics. Please review our comments policy if you have any concerns on posting a comment below.

About the authors

__Peter van den Besselaar__ is professor of organisation sciences at the Vrije Universiteit Amsterdam, the Netherlands, Faculty of Social Sciences, and research director at Teresa Mom Consultancy, Amsterdam, the Netherlands. His research focuses on the organisation, governance, and dynamics of science. In recent years, he has been engaged in an interdisciplinary team developing the SMS platform, an infrastructure for integration and enrichment of heterogeneous data, and in the GendERC project, studying gender bias in grant allocation.

__Ulf Sandström__ is Senior Lecturer at Linköping University and affiliated researcher at KTH Royal Institute of Technology in Stockholm. Currently he is working with the Gedii-team studying gender diversity in research teams. He focuses on questions related to research policy, often with bibliometric methods. Some of his latest reports and articles can be found at the forskningspolitik website.