TRANSLATIONAL TWISTS AND TURNS: SCIENCE AS A SOCIO-ECONOMIC ENDEAVOR

PROCEEDINGS OF STI 2013 BERLIN

18TH INTERNATIONAL CONFERENCE ON SCIENCE AND TECHNOLOGY INDICATORS

iFQ
Institute for
Research Information
and Quality Assurance

ENID
European Network of
Indicators Designers

Published for the 18th International Conference on Science and Technology Indicators

“Translational twists and turns: Science as a socio-economic endeavor”

organized under the auspice of the European Network of Indicator Designers (ENID)
hosted by the Institute for Research Information and Quality Assurance (iFQ)
Schützenstraße 6a
10117 Berlin
Germany
Phone +49(0)30-206 41 77-0
Fax +49(0)30-206 41 77-99
E-Mail info@forschungsinfo.de
Web www.forschungsinfo.de (German)
www.research-information.de (English)

ISSN 1864-2799

September 2013

Translational twists and turns: Science as a socio-economic endeavor

Proceedings of STI 2013 Berlin

18th International Conference on Science
and Technology Indicators
Berlin, Germany | September 4 – 6, 2013

Edited by Sybille Hinze / André Lottmann

More competition, better science? The predictive validity of grant selection

Peter van den Besselaar p.a.a.vanden.besselaar@vu.nl / Department of Organization Science &
Network Institute, VU University Amsterdam, De Boelelaan 1081,
Amsterdam, 1081 HV (Netherlands)

Abstract

Does career grant competition result in the selection of the best young talents? In this paper, the predictive validity of grant decision-making is investigated, using a sample of 250 early career grant applications in three social science fields. We measure output and impact of the applicants about nine years after the application to find out whether the selected researchers perform ex post better than the non-successful ones.

In this first test, we find that predictive validity varies sometimes is high (psychology) but sometimes low (economics), when comparing grantees with *all* non-successful applicants. However, when comparing grantees with the *best* performing non-successful applicants, predictive validity varies between low (psychology) to negative (economics). Furthermore, our findings suggest that predictive validity decreases with increasing competition, which has theoretical implications for the understanding of how panels work, and for how scholarly talent is (or cannot) recognized. The analysis is based on publications in social science journals only. In a following step, we will extend the analysis to publications in science journals.

Introduction

An important question about peer and panel review is the predictive validity: does the post-performance of selected researchers legitimize their selection: do they outperform those that were not selected? Unfortunately, data to investigate this are scarce. Not surprisingly, a recent review of research on peer review (Bornmann 2011) could only identify six studies on the predictive validity of grant peer review (Armstrong 1997; Bornmann et al, 2005, 2008, 2010; Hornbostel et al, 2009; Melin & Rickard 2006; Van den Besselaar et al 2009). Recently, a few other studies have been published, indicating the growing interest in the subject (Campbell et al, 2010; Li et al, 2010, Neufeld & von Ins 2011; van Leeuwen & Moed 2012).

Some of the available studies compare successful applicants with all unsuccessful applicants. Generally, the former are reported *in average* to outperform the latter. This holds for past performance (before the grant application) and post performance (a few years after the application). Despite this positive relation, not all selected applicants of course score higher than all rejected ones. A significant number of false positives and false negatives is reported, implying that about one third to two third of all decisions can be considered wrong if one accepts the deployed performance criteria as meaningful.

Several of these studies compared successful applicants with (an equally large set of) the *best performing non-successful* applicants. Some found a higher post-performance for the grantees, despite the fact that this was not the case for past performance. This suggests would that grant decisions may not select the best, but produce the best through providing resources. Other studies found no differences between past and post performance between the two groups.

This study builds on earlier work (Van den Besselaar et al 2009; Bornmann et al 2010). Here we go beyond our own and other earlier studies in the following ways: (i) Mid career researchers and the advanced researchers do have a variety of funding possibilities. One cannot control for that, as information lacks about applications elsewhere, e.g., the ERC career grants. However, this does not hold for early career researchers, who have mainly one funding source available. By restricting the analysis in this paper to the early career program (here: the Dutch VENI program), we most probably have disjoint sets of successful and unsuccessful applicants, not influenced by other grants. (ii) Our previous study only included a rather short period of post performance, between two and four years after the *start* of the project. As the time between having results and having those published is fairly long in the social sciences, and citing that work takes some more years, our earlier citation data (early February 2007) hardly measure real post performance. (iii) Comparisons are generally made in terms of averages. As distributions are skewed, we switch here to non-parametric statistics. (iv) The context of decision-making is taken into account, as e.g., the way the decision-making is organized, influences the process and its outcomes (Langfeld 2001; Van Arensbergen et al 2012; Van den Besselaar et al 2013); (v) data collection is improved, such as better disambiguation of authors.

Data and method

The data consist of 400 early career researchers in the social sciences, including name, age, university, field and discipline, reviewer scores, and panel decision. The applicants obtained their PhD between 2000 and 2002, and grants were obtained in the period 2003–2005. Post performance of these early career researchers is measured by all publications between 2000 and 2012, and citations received at 31-12-2012.

Post performance is defined in terms of productivity (publications), overall impact (citations), and average impact (citations per publication). Similar to the methodology used in an earlier study (Van den Besselaar et al 2009), the analysis in this paper is restricted to publications indexed in the Social Sciences Citation Index (WoS-SocSCI). Although this will underestimate output of those researchers who also publish in e.g., mathematical and statistical journals (not uncommon for economists), or life science journals (not uncommon for psychologists), we expected that this does not influence the comparison at group level. Given the size of the sample used in the 2009 study, this underestimation was expected to be equal for the successful applicants, the non-successful applicants and the best-performing non-successful applicants. However, we are currently testing this. Finally, data for this analysis were collected manually[i], in order to improve identification and disambiguation of authors.

The analysis was done for applicants within economics, psychology, and educational and behavioral sciences. In these fields publishing in international journals has become the normal practice, possibly with the exception of educational research. Other social science fields are either insufficiently covered by the WoS (anthropology, sociology, law), or too small in our sample for the current analysis (geography, communication science, organization science). Competition (success rate) is rather different between the three fields under study (Table 1).

Table 1. The sample[ii]

	Applicants	**Success rate**
Psychology	95	28.4%
Educational and behavioral research	48	20.1%
Economics	101	12.9%
Total	**244**	**20.5%**

We will compare the publications, citations, and citations per publication of successful applicants (S), the equally large set of best performing rejected applicants (BPNS)[iii], and all rejected applicants (NS). Earlier studies were generally based on comparing means. However, as distributions are non-normal and outliers cannot be discarded, one needs to turn to non-parametric statistics. We will test whether the medians between the relevant groups are different and whether the performance distributions of these groups are different (Mann-Whitney; sample median test). Based in this, we introduce a measure for predictive validity.

We will test the following four null hypotheses for each of the three disciplines:

There is no difference between

(1) *median post performance* (publications, citations, citations per publication) of the successful applicants (S) and the non-successful applicants (NS)
(2) *post performance distributions* (of publications, citations, citations per publication) of S and NS
(3) *median post performance* of S and the best performing non-successful applicants (BPNS)
(4) *post performance distributions* of S and BPNS

Testing these hypotheses will enable us to find out whether the successful applicants outperform the others by the end of 2012, implying that they either were better when applying, or have become the better through the resources and possibilities a career grant offers. An earlier analysis suggested that the selected applicants did not (ex ante) perform better than the best performing non-selected researchers (Van den Besselaar et al 2009). Are the grantees better now? Finally, by comparing the findings between the three disciplines, we may find out as whether the level of competition influences the predictive validity.

Findings

The hypotheses about the medians and the distributions of *post performance* of the three groups (S, BPNS, NS) within the three disciplines are tested, using the SocSCI publications. Table 2 shows the first results. If we compare post performance of successful applicants with *all* unsuccessful ones, we do not find performance differences within economics. We do find the granted applicants having a better post performance distribution than the non-successful applicants in psychology. Within educational and behavioral research, the finding is mixed: for some indicators, grantees' distribution is above the others' distributions, for other indicators they do not.

And what when we compare the grantees with the *best performing* non-successful applicants? Within economics, the BPNS have a better post performance distribution than the successful applicants. In psychology, the picture is mixed: according to two indicators, the grantees do better, but the four other indicators show no difference. Within educational and behavioral research, no differences between the groups are found.

In order to draw conclusions from these findings, we calculate a *predictive validity score* (PVS), which averages the predictive validity of the different indicators used. It is calculated in the following way, using applications in economics as example. We have six comparisons (median and distribution of P, C, and C/P) between the groups. When a comparison supports predictive validity, it gets a score of 1. When a comparison contradicts predictive validity, it scores -1. And if there is no difference, the score is 0. The average of the scores is used: for publications in economics (granted versus all non-granted), we found no difference between the two groups, so the score is (6*0)/6 = 0. For publications in economics (granted versus best non granted), we found for two comparisons no differences, and in four we found that the non-granted performed better. Here the predictive validity score is ((4*-1)+(2*0))/6 = -.67. The PVS-scores are in the right column of table 4. We find a rather low predictive validity score for education and an even lower and negative for economics. Psychology has a high PVS if we compare the granted applicants with all others, but in case the comparison with the best performing others, the predictive validity score drops to 0.33.

Table 2. Success by post performance and discipline – some provisional results

Null hypothesis: no differences between:			Medians*	Distributions*	PVS**
S versus NS	economics N = 101	Publications (P)	Retain	Retain	0/6 = 0
		Citations (C)	Retain	Retain	
		Citations/Publication (C/P)	Retain	Retain	
	Psychology N=95	Publications	S > NS	S > NS	6/6 = 1
		Citations	S > NS	S > NS	
		Citations/Publication	S > NS	S > NS	
	Education N=48	Publications	Retain	S > NS	2/6 = 0.33
		Citations	Retain	S > NS	
		Citations/Publication	Retain	Retain	
S versus BPNS	economics N=26	Publications	Retain	Retain	-4/6 = -0.67
		Citations	BPNS > S	BPNS > S	
		Citations/Publication	BPNS > S	BPNS > S	
	Psychology N=54	Publications	S > BPNS	Retain	2/6 = .33
		Citations	Retain	Retain	
		Citations/Publication	S > BPNS	Retain	
	Education N=20	Publications	Retain	Retain	0/6 = 0
		Citations	Retain	Retain	
		Citations/Publication	Retain	Retain	

S = successful applicants; BPNS = best performing non=-successful applicants; NS= non-successful applicants; N= number applicants

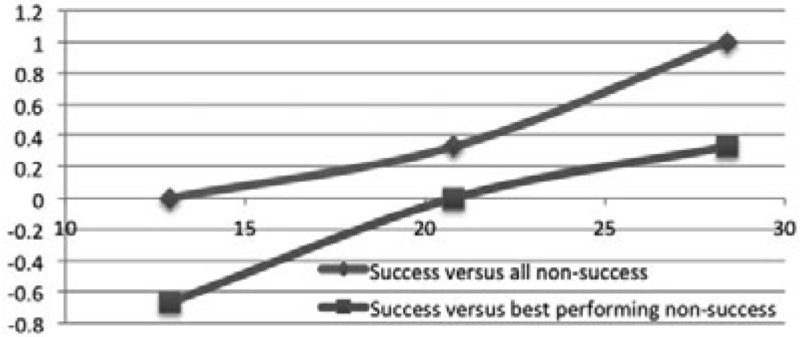
* Significance level = 0.10

** Predictive validity score (explained above in text)

Context and predictive validity – an initial analysis

In Fig. 1 gives the six predictive validity scores for the three fields, for S versus NS and for S versus BPNS by success rate. As this is a small sample, and the analysis does not include publications in non-social science journals yet, we cannot draw any definitive conclusions. Nevertheless, figure 1 suggests a positive relation between predictive validity score and success rate, confirming our expectation.

Figure 1. Predictive validity index score (Y-axis) by Success rate in percentages (X-axis)



Why would this be the case? When selecting scientific talent, panel members are influenced by scientific norms, and also by interests. What dominates may be dependent on the scarcity or abundance of resources. If resources are scarce and success rates very low, representing (disciplinary) interests may dominate. E.g., if there are not enough grants for every specialty, panelist may have highest priority to secure a grant for their own specialty over selecting the best applicants. And if selection is more based on interests than on the quality of the applicant and the proposal, predictive validity is expected to be lower. Or to put it differently, strong competition among applicants may lead to competition within panels; whereas lower competition among applicants enables collaboration in panels to select the best set of applicants. This suggests that although competition drives science, too much competition may destroy the normative order that is necessary for a good functioning science system.

Also another common belief has to be questioned. Interviews among panel members show that they are convinced that real talents are easily recognized. Within the larger sub-top, it becomes more difficult to differentiate between those that should be funded and those that should not (Van den Besselaar et al 2013). If that would be the case, the predictive validity would be better if only a few applicants have to be selected (the real top) than if a larger selection has to be made. Our findings, however, suggests exactly the opposite. The 'evidence' of talent may be a mere myth.

Conclusions and next steps

This analysis suggests that predictive validity is only high when (i) comparing grantees with all non-grantees, and when at the same time (ii) success rate is rather good. For the other situations (comparison with BPNS applicants; low success rate), predictive validity is low. We suggested some possible theoretical explanations above. However, further empirical analysis is needed too. The following next steps are on the agenda:

(1) Extending the performance measures from only SocSCI to also SCI data, to cover the output and impact more comprehensible;

(2) Testing the hypotheses in other fields;

(3) Exploring different indicators for performance. Are the (common) indicators, also used in this paper, valid for measuring post performance? As argued elsewhere, success may be related not so much to the common output and impact measures, but more to proper indicators of scholarly quality, reflecting independence (Van den Besselaar et al 2012), and scientific innovation. It is worthwhile, theoretically and from a policy perspective, to explore this further.

References

Armstrong PW, Caverson MM, Adams L, Taylor M, Olley PM (1997). Evaluation of the Heart and Stroke Foundation of Canada Research Scholarship Program: research productivity and impact. *Canadian Journal of Cardiology* 13, May, 507–16.

Bornmann L (2011). Scientific peer review. *An Rev Inf Sci Tech* 45, 199–245.

Bornmann L & Daniel H-D (2005). Committee peer review at an international research foundation: predictive validity and fairness of selection decisions on post-graduate fellowship applications. *Research Evaluation* 14, 15–20.

Bornmann L, Wallon G, Ledin A (2008) Does the Committee Peer Review Select the Best Applicants for Funding? An Investigation of the Selection Process for Two European Molecular Biology Organization Programmes. *PLoS ONE* 3 (10): e3480.

Bornmann, L, Leydesdorff L, Van den Besselaar P (2010). A Meta-evaluation of Scientific Research Proposals: Different Ways of Comparing Rejected to Awarded Applications. *Journal of Informetrics* 4 (3) 211–220

Campbell D, Picard-Aitken M, Cote G, Caruso J, Valentim R, Edmonds S, Archambault, E (2010) Bibliometrics as a performance measurement tool for research evaluation: the case of research funded by the National Cancer Institute of Canada. *Am J Eval* 31, 66–83.

Hornbostel S, Böhmer S, Klingsporn B, Neufeld J, von Ins M (2009). Funding of young scientist and scientific excellence. *Scientometrics* 79, 171–190

Langfeldt L (2001). Decision making constraints and processes of grant peer review, and their effect on review outcome. *Social Studies of Science*, 31, 820–841

Li J., Sanderson M, Willett P, Norris M & Oppenheim C (2010). Ranking of library and information science researchers: Comparison of data sources for correlating citation data, and expert judgments. *Journal of Informetrics* 4 (4) 554–563.

Melin G, Rickard D (2006). The top eight percent: Development of approved and rejected applicants for a prestigious grant in Sweden, *Science and Public Policy* 33, 702–712

Neufeld J & von Ins M (2011). Informed peer review and uninformed bibliometrics? *Research Evaluation* 20, 31–46.

Van Arensbergen P & Van den Besselaar P (2012). The selection of scientific talent in the allocation of research grants, *Higher Education Policy* 25, 381–405

Van Arensbergen P, Van der Weijden, Van den Besselaar P (2013), The notion of talent, What are the talents we are looking for in science? In: *STI 2013 Proceedings – this volume*.

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

Van den Besselaar P & Van Arensbergen P (2013). Talent selection and the funding of research. Academic talent selection in grant review panels. *Higher Education Policy* 26.

Van den Besselaar P, Sandström U, Van der Weijden I (2012) The independence indicator. In Archambault E, et al, eds., *Proceedings STI 2012*. Montreal: OST & Science Metrix.

Van Leeuwen TN & Moed HF (2012). Funding decisions, peer review, and scientific excellence in physical sciences, chemistry, and geosciences. *Research Evaluation* 21, 189–198.

[i] In contrast, the 2009 study with Leydesdorff is based on automatic mapping between WoS data and application data, using family name and first initial. Spelling variants leading to under- and overestimation were accepted, because of the large number of cases. However some other issues have come up since. (i) The first initial was sometimes taken from the title (PhD, Dr, MSc, LLM, etc.). (ii) Some applicants use different initials in applications compared to publications. Re-analyzing after (partly) correcting (i) and (ii) suggests that this does not influence outcomes of the 2009 study.

[ii] Sixteen applications were removed because of missing data.

[iii] Best performing rejected applicants is defined in terms of total citations received.