Research Policy 44 (2015) 467–482

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Does massive funding support of researchers work?: Evaluating the impact of the South African research chair funding initiative☆

J.W. Fedderke[a,b,c,d,*], M. Goldschmidt[a]

[a] Pennsylvania State University, USA
[b] Economic Research Southern Africa, South Africa
[c] South African Reserve Bank, South Africa
[d] University of the Witwatersrand, South Africa

ARTICLE INFO

Article history:
Received 15 October 2013
Received in revised form
24 September 2014
Accepted 25 September 2014
Available online 31 October 2014

Keywords:
Public research funding
Impact research
Output research
Impact bibliometric measures of research performance
Productivity of research
Funding allocation for research

ABSTRACT

In this study we evaluate whether a substantial increase in public funding to researchers is associated with a material difference in their productivity. We compare performance measures of researchers who were granted substantial funding against researchers with similar scholarly standing who did not receive such funding. We find that substantial funding is associated with raised researcher performance – though the increase is moderate, is strongly conditional on the quality of the researcher who receives the funding, and is greater in some disciplines than others. Moreover the cost per additional unit of output is such as to raise questions about the usefulness of the funding model. The implication is that public research funding will be more effective in raising research output where selectivity of recipients of funding is strongly conditional on the established track record of researchers.

1. Introduction

Strategic funding of research and development by public entities continues to be viewed as important to the ability of both business and other types of organization to innovate.[1] This makes funding for science of perennial concern, since the financial requirements of projects that carry scientific merit outstrips both public and private budgetary provision, and faces significant competition from other social spending priorities. In such a context, it is important to understand which funding mechanisms of science are effective, and which are not.

This raises a fundamental management issue. Like many other investment projects, investment in research and development may require substantial up-front outlays on the promise of future success and returns in the form of increased knowledge. Yet investment in innovation represents commitment of resources to projects with an uncertain rate of return in knowledge. While any investment faces uncertainty,[2] where investment is in knowledge creation, uncertainty is magnified.[3] The management difficulty is that if research funding is organized so as to front-load the funding commitment to the researcher, the feasibility of relying on incentive mechanisms that reward research based on the delivery of successful innovations becomes limited. As such, reliance on standard reward structures to incentivize productivity becomes severely constrained.

In a market setting some recent research provides insights into efficient approaches to resource allocation. For instance, Klingebiel and Rammer (2013) in a study of innovation drawing on firm level evidence present results that support funding allocation

☆ The research support of Economic Research Southern Africa is gratefully acknowledged. The views expressed in this article are those of the authors alone, and do not necessarily reflect those of the South African Reserve Bank. Useful comments by anonymous referees, and editorial guidance on an earlier version of the paper significantly improved its contents. Responsibility for remaining errors remains ours alone. Research assistance from Wen Chen, Yang Liu, and Sandeep Regmi is gratefully acknowledged.

* Corresponding author at: Pennsylvania State University, USA. Tel.: +1 814 777 4777.
E-mail address: jwf15@psu.edu (J.W. Fedderke).

[1] See for instance the discussions in Fagerberg (1994), Mowery and Rosenberg (1989), Nelson and Wright (1992) and Nelson (1992, 1996).

[2] See the seminal discussion in Dixit and Pindyck (1994).

[3] The high uncertainty and hence risk attaching to innovation and research and development is the subject of a substantial literature. See for instance the introductions in Mokyr (2002) and Rosenberg (1994).

http://dx.doi.org/10.1016/j.respol.2014.09.009
0048-7333/

across a broad range of projects as more successful than a more focussed resource-intensive allocation, since winners are difficult to predict.[4] The effectiveness of the broad-based funding allocation is found to be enhanced if coupled to monitoring tied to later-stage selection of successful innovations.[5]

In a public funding context, there are additional difficulties. Price mechanisms provide a disciplining device on market-based agents such as firms that are not present for public funding agencies devoted to the management of research. Firms can signal their prospects of successful innovation in capital markets through the return on financing that they are prepared to offer. Researchers applying for funding from public agencies have no equivalent price mechanism at their disposal. What is more, there is no reason to believe that the standard problems of moral hazard and adverse selection that characterize capital markets, would be absent from the signalling that researchers engage in when submitting funding proposals to public research funding agencies.

Problems associated with public research funding are not restricted to the difficulty of correctly assessing noisy signals received from applicants for funding. A number of empirical research findings show that public funding agency evaluations of researchers are often only weakly tied to the objective output and impact performance of researchers, and that funding allocations are similarly weakly correlated with research output and impact measures, in part because of inherent conservative risk-averse biases in public funding agencies that inhibit innovation.[6] The fundamental difficulty is that absent objective performance based output measures as the criterion of merit and funding allocation, there is an irreducible subjective element in the assessment process.[7] An alternative approach that has been suggested to circumvent these difficulties faced by public funding agencies, is the explicit use of performance-based output measures. Structuring recognition on objective output measures reduces conservative and subjective biases, and reduces the risk of adverse selection and moral hazard problems in the distribution of resources.[8]

Such complex and multi-faceted difficulties confronting public research management make it difficult to isolate the relative empirical significance of the various aspects of the challenges. In this paper we make use of a policy intervention, that initiated a very resource-intensive investment in research through a public agency targeted at a small group of researchers. We track the research output and impact of the group of researchers who receive substantial public funding in the form of a research chair, against a set of control groups of researchers of equivalent scholarly standing, who do not receive such funding.

Our findings suggest that while the performance of research chair holders does improve after the funding allocation, the improvement is moderate, strongly conditional on the research standing of the awardee at the time of the award, and differentiated across disciplines. Moreover, per additional unit of research output, research chairs prove very expensive.

While we present a range of policy inferences, there are two core implications that we emphasize. First, funding allocations should be selective in the sense of being responsive to past research performance. Our results show that highly rated researchers who receive funding show stronger productivity increases than those with low ratings, suggesting that productivity increases will be greatest when focussed on researchers with strong track records. Second, since the cost per additional unit of research output rises dramatically even for the most productive research chair holders relative to equivalent researchers without chairs, suggests a strongly decreasing marginal productivity of increases in research funding. The policy inference we draw is that smaller funding grants to more researchers with strong track records is therefore likely more effective than massive funding concentrated on a few researchers alone. Conditionally strictly on the demonstration of adequate performance increases after the initial funding allocation, funding can always subsequently be scaled upward, and concentrated on successful research initiatives.

Central to such an approach is the need for revealed productivity to be transparently and objectively monitored. The growing number of objective bibliometric measures, whose collection is greatly facilitated by the growth in information technology, offers an immediate means of doing so.

We proceed as follows. In Section 2 we provide a precise statement of the research question and explain the associated modeling strategy. Section 3 details data sources, Section 4 presents results, while Section 5 concludes.

2. The research question and methodology

In this study we empirically examine whether a substantial increase in public funding allocations to researchers is associated with a material difference in their productivity. To do so, we compare the scholarly performance, in terms of both output and impact as measured by objective bibliometric measures under the Thomson ISI Web of Science citations database, of a body of researchers who were granted substantial research funding, against the performance of a body of researchers of similar inherent scholarly standing, who did not receive such funding.

Our focus is on South African data. The reason for this is that in 2008–2009, the National Research Foundation of South Africa (NRF) awarded a total of 80 research chairs, each of which was endowed with substantial research funding (approximately US$300,000 per annum) guaranteed over a period of 5 years, renewable for up to 15 years.[9] The stated goal of the NRF research chairs is to improve South Africa's competitiveness in the international knowledge economy by expanding scientific research and innovation

[4] See also Leiponen and Helfat (2010, 2011). The literature has also considered a range of additional determinants such as openness – see Aghion et al. (2013a,b) at an aggregate level and Laursen and Salter (2006) at micro level – and other features of strategic management and organizational structure – see Aghion et al. (2013b), Cassiman and Veugelers (2006), Garriga et al. (2013), Leiblein and Madsen (2009) and Li and Atuahene-Gima (2001). Approaches that tie inputs to research and development success are the subject of a literature in their own right – see for instance Crépon et al. (1998), Mairesse and Mohnen (2002) and Van Reenen (2011).

[5] Klingebiel and Rammer (2013) also discuss the dangers associated with broad-based allocation mechanisms – particularly the dissipation of resources, lack of strategic focus, and diminished incentives.

[6] On the poor correlation of evaluation and funding with objective performance see Fedderke (2013) and Grimpe (2012). On the conservative biases in public funding agencies see Braun (1998).

[7] See the discussion in Moxhan and Anderson (1992) and Horrobin (1990).

[8] See for instance the discussion in Abramo et al. (2009), Butler (2003) and Hicks (2012). There are also dangers to the approach. Butler (2003) indicates that it may favour quantity over quality. Hicks (2012) notes that performance-based reward structures focus specifically on excellence. Van Raan (2005) reports concerns regarding the lack of homogeneity in coverage of different scientific areas, and sensitivity to starting data bases. Korrevaar and Moed (1996) point to the significance of potential disciplinary differences. However, these are limitations that can be overcome by the use of multiple measures of performance, and careful bottom-up construction of data. See the discussions in Abramo and D'Angelo (2007), Aksnes and Taxt (2004), Martin (1996), Oppenheim (1997) and Rinia et al. (1998). Ederer and Manso (2012) present experimental evidence demonstrating that reward for performance does not carry the disincentive effects feared in applied psychology, conditional on the correct design of the reward structure (mitigation of risk is important).

[9] See http://www.nrf.ac.za/sarchi/index.stm for a full description of the initiative. In 2012 the NRF awarded an additional set of chairs. These are not included in our analysis, since not enough time has passed to assess the impact of the new chairs. Not all research chairs necessarily received the full funding quota available – but even the minimum funding granted ($150,000) was considerably larger than that granted non-chair researchers.

capacity and increasing the number of world-class researchers in South Africa.[10] The 2008–2009 round of research chair awards was the first set of NRF research chairs granted. It marks the first instance in which this funding was available to researchers and the effectiveness of the initiative can be explored.

The NRF chair selection process has two phases: During the first phase the universities that will host the chair apply for the funding, a panel reviews the proposals and makes recommendations based on the compatibility between the proposed chair and both the capabilities and strategic alignment of the host university. During the second phase, the universities submit proposals to nominate candidates to be chair holders. The applications are reviewed first by a group of peers and then a panel in order to approve the funding and the nominee. It is important to highlight that the NRF states that they fund candidates, rendering candidate credentials key to the selection process.[11]

In addition to the research chair initiative, the NRF also operates a peer review mechanism that is designed to rate scholars in all disciplines. The review mechanism is independent of the research chair mechanism, and rates both research chair holders and those scholars that do not hold a research chair against in the same manner. In the peer review mechanism, scholars apply for a NRF rating. A subject specific committee with 6–10 peer reviewers evaluate the applicants based on output and impact of their research over the last seven years as well as the standing of the applicants among their peers. The peer review can issue in four possible ratings. An A-rating is held to apply to researchers who are unequivocally recognized by their peers as leading international scholars in their field for the high quality and impact of their recent research outputs. The B-rating is awarded to researchers who enjoy considerable international recognition by their peers for the high quality and impact of their recent research outputs. A C-rating applies to researchers who have a sustained record of research productivity, while Y-ratings are awarded to young researchers.

Crucially for our purposes, researchers under the peer review mechanism, while being rated on the same metric as the research chair holders, receive considerably less research funding ($10,000 and $8000 per annum for the very highest ratings attainable). This allows the performance of researchers with high funding allocations (the NRF chairs) to be compared to those with much lower funding allocations, who have identical research standing in terms of the funding body's own metrics.

But is the NRF peer review mechanism reliable enough to render the researchers comparable? The findings reported in Fedderke (2013) show that absolute publications output as well as impact raise the probability of any specific NRF rating, the level of the NRF rating of a scholar, as well as the probability of being awarded a NRF chair.[12] These findings lend support to the functionality of the NRF peer review process in rendering similarly rated researchers comparable.

Nonetheless, since the NRF ratings and chair granting mechanisms rely on peer review, they could be subject to potential sources of bias. Consistent with this possibility Fedderke (2013) found the variance of objective performance of NRF chair holders to be large. Some scholars who have received NRF research chairs have levels of research output and impact similar to scholars located at the lowest end of the performance distribution of all scholars who are rated by the NRF. This creates a contradiction with the objective that NRF chairs attract world class scholars, and raises prospects of bias in the allocation of chairs. In addition, there are strong disciplinary differences in terms of the impact that improved performance (under the *h*-index measure) has on the probability of getting a NRF Chair. Simple reliance on the NRF peer review mechanism, without suitable tests for the robustness of results derived under it, are thus unlikely to be conclusive.

For these reasons, this paper uses two distinct methodological designs. The first methodological design utilizes the NRF peer review based rating mechanism to compare the performance of NRF research chair holders against the performance of control groups of scholars of comparable standing under the peer evaluation mechanism, who did not receive the funding under the research chair initiative.[13] NRF research chairs are targeted at researchers with international academic leadership status. For this reason this study focusses on only two NRF rating classes, A- and B-rated scholars since these are the groupings of researchers deemed of internationally competitive standing, the stated target of the research chair initiative. The funding allocation by the NRF to A- and B-rated researchers is approximately US$10,000 and US$8000 per annum respectively. Given the $300,000 award to NRF research chair holders, this implies that NRF research chairs receive funding grants 30 times as large as A-rated researchers, and approximately 38 times as large as B-rated researchers. In order to test the success of the funding intervention in generating higher research productivity, we compare the scholarly output and impact of researchers who did receive the funding associated with a NRF research chair, with scholars of comparable standing who did not. We begin by recording the standing of three groups of researchers (NRF research chairs, A-rated researchers, B-rated researchers) at the time the NRF research chairs were awarded in 2009, across a range of bibliometric measures of absolute output, of scholarly impact, and in terms of the *h*-index. We then record the change in the performance of the three groups of researchers from 2009 to 2012, across the same range of bibliometric measures, in order to establish whether there is an appreciable difference in performance between the NRF research chair incumbents, and A- and B-rated researchers who did not receive the funding associated with a NRF research chair.

The second methodological design utilizes propensity score matching against objective bibliometric measures of scholarly performance. We first record the standings of the sample of researchers at the time NRF research chairs were awarded in 2009 across a range of bibliometric measures of absolute output, of scholarly impact, and in terms of a composite measure of absolute output and of impact. Using a logit regression model, we calculate the probability of obtaining a NRF chair award based on three observable characteristics at the time of the award (2009): (1) the number of publications, (2) the number of citations and (3) the researchers' disciplinary field of study. We compute propensities both on publications and citations only, and by using all three characteristics, estimate the probability of receiving an award given observable characteristics for each scholar in the sample, and divide our sample into three control groups based on the probability of receiving a NRF chair (low, medium and high probability). This methodology allows us to create a set of control groups against which the performance of highly funded researchers can be compared, devoid of the potential subjective bias associated with peer review mechanisms. Symmetrically with the first methodology, we then record

[10] http://hicd.nrf.ac.za/?q=node/16.

[11] http://hicd.nrf.ac.za/?q=node/27.

[12] Fedderke (2013) found that of the 80 chairs listed by the NRF in 2008–2009, 71% of researchers were rated under the NRF peer review system, leaving 29% of the NRF chair holders unrated. Of the rated NRF chairs, 10% held an A-rating, 36% a B-rating, 23% a C-rating, and 3% a Y-rating.

[13] Peer review is itself subject to strengths and weaknesses – see the review in Bornmann (2011). For our purpose we only require that the NRF chairs are compared with comparable researchers, for which the NRF peer review process which considers all researchers against the same standards provides a means of selection. If this still leaves doubt regarding comparability, we also use the propensity score matching methodology.

the change in the performance of the three groups of researchers from 2009 to 2012, across the range of bibliometric measures, in order to establish whether there is an appreciable difference in performance between the NRF research chair incumbents and each of the three control groups (low, medium and high probability of receiving funding).

As the empirical results will demonstrate, both methodological designs reach the same inferences and conclusions, lending robustness to our findings.

The advantage of the research design is as follows. There is a specific temporal point at which a clearly identifiable set of researchers received a substantial increase in funding toward their research. Since the NRF operates an independent peer review mechanism of researchers, the scholarly standing of the NRF chair incumbents is readily identifiable. There is an equally clearly identifiable control group of researchers available, who are equivalently rated by the NRF, but who were not granted the funding associated with a research chair. We have a set of objective bibliometric measures of performance in terms of which the performance of researchers can be measured. Finally, since the NRF funding body covers all academic disciplines in South Africa, the study is not simply a reflection of performance in one area of intellectual endeavour, but carries general implications for the performance of the academy as a whole. It also allows for cross-disciplinary comparisons in the results we derive.

But why the South African data, especially since the South African research chair initiative is modelled on the Canadian research chair initiative – see Neuman (2003) and Polster (2002)? The reason for choosing the African data is that Canada does not allow for the comparison of the research chair's performance against a comparable peer review selected control group, since it does not have a rating system of researchers with near universal coverage as does South Africa. The Canadian research chairs initiative was created in 2000 with a goal of attracting and retaining "accomplished and promising minds". The program has a two-tier structure. Tier one (CAN$200,000 annually) is tenable for seven years, can be renewed indefinitely and is granted to outstanding researchers. Tier two (CAN$100,000 annually) is held for five years, is renewable once and is granted to exceptional emerging researchers.[14] As with the South African research chair initiative, the Canadian research chairs are first allocated to universities and then to particular researchers who are nominated by the universities and evaluated by a peer review panel. The universities' share of research chairs depends on past success in previous granting competitions, Polster (2002). An important difference between the NRF and Canada chairs therefore, is that officially the NRF chairs were aimed directly at the attraction of leading researchers, while the Canadian program also had a developmental component for newly emerging researchers deemed worthy of stronger support.

2.1. Our hypotheses

Our interest lies in three issues. Is a large increase in research funding associated with increased research productivity? Is the increased productivity (if present) conditional on the quality of the researcher who receives the funding? Is the increased productivity (if present) differentiated across disciplines?

The paper therefore examines three separate hypotheses.

Hypothesis 1 *(H1).* Scholars who received the higher level of funding associated with a NRF chair award have a significantly higher performance as measured by output and impact than researchers of comparable standing who did not receive a research chair.

We examine H1 by estimating:

$$\textit{Peer Review Mechanism}: \quad \Delta M_i = \alpha A_i + \beta B_i + \eta N_i + \varepsilon_i \tag{1}$$

$$\textit{Propensity Score Design}: \quad \Delta M_i = \alpha H_i + \beta U_i + \gamma L_i + \eta N_i + \varepsilon_i \tag{2}$$

where ΔM_i denotes the change in the bibliometric index of interest for researcher i, A_i and B_i denotes a categorical variable for an A- and B-rating respectively for the ith researcher not holding a NRF chair, while N_i denotes a categorical variable for a NRF chair holder. Under the propensity score approach H_i denotes the control group of scholars with high probability of obtaining a NRF chair, U_i denotes the control group of scholars with medium probability of obtaining a NRF chair, L_i denotes the control group of scholars with low probability of obtaining a NRF chair. The H_i, U_i, and L_i control groups do not include NRF chairs holders. We denote the error by ε_i.

Our question is whether $\eta \neq 0$, and $\eta > \alpha$, $\eta > \beta$, $\eta > \gamma$.

Hypothesis 2 *(H2).* The higher the scholarly standing of the recipient of the funding associated with NRF chairs in 2009, the higher the increase in performance as measured by output and impact.

To explore H2, i.e. whether the change in researcher performance is conditional on the scholarly standing of the recipient of the chair in 2009, under the peer review mechanism we estimate:

$$\textit{Peer Review Mechanism}: \quad \Delta M_i = \alpha A_i + \beta B_i + \sum_{k=1}^{3} \delta_k R_k N_i + \varepsilon_i \tag{3}$$

$$\textit{Propensity Score Design}: \quad \Delta M_i = \alpha H_i + \beta U_i + \gamma L_i + \sum_{j=1}^{3} \delta_j C_j N_i + \varepsilon_i \tag{4}$$

where notation is defined as before and R_k is a categorical variable for three possible ratings that NRF recipients held at the point where the chair was awarded in 2009, viz., an A-rating, a B-rating, or any rating other than A or B. C_j is a categorical variable for the three propensities of receiving a NRF chair, viz. H_i, U_i, and L_i.

Our question is whether the δ_k, δ_j, differ across the researcher ratings or propensities, and from the α, β, γ parameters.

Hypothesis 3 *(H3).* Research productivity of NRF research chairs is conditional on the discipline of the researcher.

To allow for disciplinary differences in terms of the productivity of the NRF research chair holders, we consider:

$$\textit{Peer Review Mechanism}: \quad \Delta M_i = \alpha A_i + \beta B_i + \sum_{j=1}^{7} \lambda_j D_j N_i + \varepsilon_i \tag{5}$$

$$\textit{Propensity Score Design}: \quad \Delta M_i = \alpha H_i + \beta U_i + \gamma L_i + \sum_{j=1}^{7} \lambda_j D_j N_i + \varepsilon_i \tag{6}$$

where D_j denotes a set of categorical variables for the seven disciplinary groupings we code in our data, the Biological, Business and Economic, Chemical, Engineering, Medical, Physical and Social sciences.

Our question is whether the λ_j, differ across disciplines.

Collectively, the three sets of specifications examining our hypotheses give insight into the production function linking

[14] See the Canadian research chairs website. http://www.chairs-chaires.gc.ca/about_us-a_notre_sujet/index-eng.aspx.

Table 1
Sample characteristics.

	Population	Sample	Percentage of population	Excluded	Records in study
NRF chair	80	80	100	4	76
NRF: A-rated	9	9	100	0	9
NRF: B-rated	30	30	100	2	28
NRF: Other rated	41	41	100	1	39
A-rated w/o NRF chair	68	68	100	1	67
B-rated w/o NRF chair	441	161	39	4	157

Table 2
Distribution of researchers by NRF rating across propensity score matching control groups.

	Low probability control		Med. probability control		High probability control	
Propensity score	P&C	P&C&D	P&C	P&C&D	P&C	P&C&D
NRF	15	15	25	23	36	38
A-rated	36	37	17	24	14	6
B-rated	49	48	58	53	50	56

P&C denotes Publications and Citations selection.
P&C&D denotes Publications and Citations and Disciplines selection.

research productivity to funding inputs. Evidence on H1, establishes whether funding recipients show differential productivity relative to our control groups ((1) and (2)), providing evidence on the return to funding on average across all research chair funding recipients. The evidence on H2 and *H3* indicates whether the return on the research chair funding is concentrated either on specific classes of researchers, either in terms of prior research track record under the NRF peer review or the propensity score mechanism ((3) and (4)), or in terms of the disciplinary home of the researcher ((5) and (6)). Thus the H2 and H3 evidence provides insight into whether the return on research chair funding is differential across different types of researcher. Since we consistently control for researchers who did not receive the research chair funding, and since we specify different types of non-research chair recipients in terms of prior research productivity, the evidence also provides insight into the raw research productivity return net of the funding granted to chair holders conditional on the quality of the researcher.

3. Data

For this study we employed two sources of data. The first was derived directly from the published list of rated scholars and research chairs of the NRF, in order to identify NRF chair holders, as well as A- and B-rated researchers. The second data source involved the compilation of bibliometric data on all identified researchers from the ISI Thomson Web of Science citations database.

From the NRF we obtained lists of all researchers that received a NRF chair in the 2008/2009 round of awards for a total of 76 records (four names were excluded due to data availability limitations); all A-rated researchers without NRF chairs, for a total of 67 records (one name was excluded due to data availability limitations); a random sample of B-rated researchers without NRF chairs, for a total of 157 records (four names were excluded due to data availability limitations). Note that while we employ the full population of A-rated researchers, we employ a 39% sample of the full population of B-rated researchers. NRF chair holders may also hold a NRF researcher rating (and typically, though not universally do so). Of the 76 NRF chairs included in the study, 9 were A-rated, 28 B-rated, and 39 either held a rating lower than A or B, or no rating at all. The fact that NRF chair holders are of diverse scholarly standing as established by the NRF peer review mechanism, allows us to examine whether NRF research chair performance is conditional on the quality of the researcher to whom the funding is given. Details of the sample are reported in Table 1.

We used the same sample of researchers to conduct the propensity score matching exercise. We create three groups of researchers based on low, medium and high probability of receiving NRF chair funding, based either only on publications and citations or on publications, citations and discipline, each group containing 100 scholars. Each of the resultant groups contains NRF chair recipients, as well as A-rated and B-rated scholars, such that the control groups are therefore distinct from those obtained under the peer review mechanism. Details of the resultant sample are reported in Table 2.[15]

Four bibliometric measures were employed to measure a researcher's standing in 2009, as well as the change in their standing over the 2009–2012 period.[16] Total Publications measures the total accumulated number of publications attributed to an author. Total citations measures the total number of citations to the work of a researcher. Average citations per Item is the Total citations count normalized on the Total Publications count. Hirsch's *h*-index is a measure which provides a composite measure both of absolute output and of the impact of the output.[17]

We employed the ISI Web of Science search engine to generate the bibliometric data for each author. Two major issues arise from the use of any search engine. The first is that each of the major alternatives (e.g. ISI, Scopus, Google Scholar) has strengths

[15] One concern that might arise here surrounds selection bias, since we might be excluding a significant body of researchers of considerable research standing, who hold neither a research chair nor an NRF rating. However, since universities and research bodies in South Africa have strong incentives (reputational, financial) to ensure that all active researchers obtain an NRF rating, and place considerable pressure on researchers to do so, virtually all active researchers in South Africa do hold an NRF rating. Hence any selection bias due to the exclusion of a significant body of active researchers that do not hold an NRF rating is likely negligible.

[16] We also considered three additional bibliometric measures: Citations without Self-citations; Total Citing Articles; Total Citing Articles without Self-citations. Since the results from these measures do not differ materially from those we report, we omit them for the sake of parsimony.

[17] The measure is well-established in the literature. A scholar has an index of *h* if *h* of his/her papers have at least *h* citations each, and the remaining papers have no more than *h* citations each. Though there are now a wide array of composite bibliometric measures available with varying properties, the *h*-index is arguably the most widely cited objective measure of scientific standing, which explains our use of the metric. For a discussion of the properties of the *h*-index, see Bornmann and Daniel (2005), Bornmann and Daniel (2007), Cronin and Meho (2006), Egghe and Rousseau (2006), Glänzel (2006), Hirsch (2005) and Van Raan (2006).

Table 3
Mean values of bibliometric indexes: raw denotes recorded ISI or PP values, Adj denotes discipline weight adjusted measures, ISI denotes ISI Web of Science results, PP denotes Publish or Perish Google Scholar based search results.

	Total publications		Total citations		Average citations				*h*-Index			
	(1) Raw		(2) Raw		(3) Raw		(4) Adj.		(5) Raw		(6) Adj.	
	ISI	PP	ISI	PP	ISI	PP	ISI	PP	ISI	PP	ISI	PP
NRF chair (All)	56.22	86.74	942.80	834.21	12.44	8.59	17.96	12.92	16.59	10.74	17.66	12.41
NRF: A-rated	83.78	105.00	1831.22	1403.11	14.57	12.69	23.55	18.09	21.78	13.44	24.57	15.50
NRF: B-rated	77.64	103.00	1330.11	1111.79	16.89	10.49	24.49	14.95	22.50	14.29	23.81	15.51
NRF: Other rated	34.49	70.85	459.72	503.64	8.76	6.28	11.98	10.28	11.15	7.56	11.66	9.47
A-rated (w/o chair)	141.64	193.96	2398.43	2017.22	12.03	7.86	14.95	10.54	24.51	17.99	24.56	18.60
B-rated (w/o chair)	59.61	93.25	719.45	602.59	9.11	5.78	13.12	9.36	14.46	10.46	15.55	12.11

Table 4
Mean values of bibliometric indexes using propensity score matching.

	Publications to 2009		Citations to 2009		Average citations to 2009		*h*-Index to 2009	
Propensity score	(1) P&C	(2) P&C&D	(3) P&C	(4) P&C&D	(5) P&C	(6) P&C&D	(7) P&C	(8) P&C&D
Low Prob w/o NRF	154.52	144.75	2119.44	2026.53	10.89	9.99	25.33	23.01
Med Prob w/o NRF	58.13	45.69	885.25	670.52	10.94	7.85	16.83	12.51
High Prob w/o NRF	21.17	48.82	423.48	802.65	7.66	12.61	7.78	16.03
Low Prob with NRF	119.20	104.87	1288.13	1120.73	9.77	9.16	24.00	21.80
Med Prob with NRF	52.88	37.74	697.24	524.43	11.14	8.05	16.64	10.61
High Prob with NRF	32.31	48.21	969.44	1125.79	14.46	16.40	13.47	18.16

P&C denotes Publications and Citations selection.
P&C&D denotes Publications and Citations and Disciplines selection.

and weaknesses.[18] In what follows, we therefore examine the sensitivity of findings to the use of Google Scholar rather than ISI to generate the bibliometric indexes. Inferences of the present study are not sensitive to the search engine employed.[19]

The second major concern regarding our data arises from the fact that the researchers in our data are drawn from very diverse disciplines, ranging from the performing and fine arts to the quantitative sciences. Evidence from the literature suggests that there is strong cross-disciplinary variation in bibliometric indices.[20] For this reason we also explore the sensitivity of our results to coding scholars in terms of the broad disciplinary fields proposed by Iglesias and Pecharromán (2007), and adjusting their bibliometric scores in the light of disciplinary weights as determined by their principal institutionally defined disciplinary affiliation. Again, inferences of the present study are not affected by the use of either the raw data, or data under the disciplinary weightings.[21]

4. Did NRF chairs demonstrate improved research productivity?

Our results cover three distinct questions. After characterizing the performance and impact of scholars at the point of the NRF research chair awards, we ask whether the performance of the NRF chairs shows significantly greater improvement than the various control groups used for this study (H1). We also ask whether the performance change of NRF research chairs is significantly associated with their research standing at the point of being granted the associated funding (H2). Finally, we question whether disciplinary differences are significantly associated with the performance of NRF research chairs (H3).

4.1. Characterizing the research performance of scholars as of 2009

In this section we explore the performance characteristics of the researchers that fall into the various categorizations employed for our study.

4.1.1. NRF peer review system

Mean values across our sample, for the standing of scholars at the time the NRF research chairs were awarded in 2009, are reported in Table 3. What is evident from the central tendency data is that the holders of NRF research chairs had absolute levels of impact (Total citations), and a joint output and impact measure (*h*-index) which lay between the A- and B-rated scholars without

[18] A non-comprehensive list of concerns with ISI-based searches are that it: does not include citations to scholarly output that has even small mistakes in its referencing; is subject to more citation noise; provides overrepresentation to English language and US and UK based journals; is biased toward citations to journal articles (as opposed to books, etc.); significantly restricts citations to non-ISI database journals; underreports citations in disciplines with long delays to publication; underreports citations in general; is sensitive to institutional subscriptions. The reliability of Google Scholar has also been questioned, on the grounds of attribution of publications to phantom authors, inclusion of non-scholarly publications, exclusion of some important scholarly journals, uneven disciplinary coverage, less comprehensive coverage of publications prior to 1990, and inconsistent accuracy. See the discussion in Archambault and Gagné (2004), Belew (2005), Bornmann et al. (2009), Bosman et al. (2006), Butler (2006), Derrick et al. (2010), Falagas et al. (2008), García-Pérez (2010), Gray et al. (2012), Harzing (2007–2008), Harzing (2008), Jacsó (2005, 2006a,b, 2010), Kousha and Thelwall (2007, 2008), Kulkarni et al. (2009), Meho and Yang (2007), Nisonger (2004), Roediger (2006), Testa (2004), and Vaughan and Shaw (2008).

[19] Given the invariance of results, we generally suppress reporting them in what follows for the sake of parsimony. Results are available form the authors upon request. For a detailed discussion of the impact of alternative search engines on the South African data, see Fedderke (2013).

[20] See the discussion in Rehn et al. (2007) and particularly Iglesias and Pecharromán (2007).

[21] Again, since results are invariant to the discipline adjustment, we generally suppress these results in what follows. Results are available form the authors upon request.

Table 5
Research trajectory in citations 2005–2009.

Citations, 2005–2009						
	Peer review mechanism					
	A-rated (w/o chair)	B-rated (w/o chair)	NRF chair (All)	NRF: A-rated	NRF: B-rated	NRF: Other rated
Trend	23.50*** (6.71)	9.01*** (0.32)	18.99*** (0.86)	35.57 (28.40)	32.51*** (5.40)	8.07*** (2.08)
Constant	110.69*** (14.85)	36.08*** (0.58)	50.34*** (2.50)	100.71* (57.02)	61.57*** (10.94)	20.25*** (4.26)
Adj-R^2	0.02	0.02	0.03	0.04	0.06	0.02
N	334	784	379	44	139	194
	Propensity score matching					
	Low Prob w/o NRF	Med Prob w/o NRF	High Prob w/o NRF	NRF: Low Prob	NRF: Med Prob	NRF: High Prob
Trend	18.44*** (4.26)	12.55*** (0.70)	6.72*** (2.04)	30.46*** (5.02)	11.89 (0.67)	19.50*** (1.51)
Constant	101.12*** (9.70)	43.03*** (1.50)	22.79*** (4.15)	55.25*** (10.13)	46.80*** (1.40)	49.34*** (4.57)
Adj-R^2	0.02	0.03	0.01	0.11	0.04	0.02
N	424	374	319	74	124	179

Figures in round parentheses denote standard errors.
* Denotes significance at 10% levels.
** Denotes significance at 5% levels.
*** Denotes significance at 1% levels.

Table 6
Change in researcher performance from 2009 to 2012: isolating citations to new work only.

Change in:	Publications 2010–2012 (1)	Citations 2010–2012 (2)	Avg. citations 2010–2012 (3)	*h*-Index 2010–2012 (4)
A-rated (w/o chair)	31.24*** (3.37)	159.30*** (25.68)	3.13*** (0.58)	0.43*** (0.13)
B-rated (w/o chair)	14.34*** (2.20)	57.07*** (16.77)	2.79*** (0.38)	0.22** (0.09)
NRF chair (All)	27.38*** (3.17)	141.63*** (24.11)	4.12*** (0.54)	0.56*** (0.12)
Adj-R^2	0.06	0.04	0.01	0.01
N	300	300	300	300
	χ^2 test of parameter equality			
A-rated (w/o chair) = NRF chair (All)	0.70 [0.40]	0.25 [0.62]	1.56 [0.21]	0.55 [0.46]
B-rated (w/o chair) = NRF chair (All)	11.43*** [0.00]	8.29*** [0.00]	4.01** [0.05]	5.46** [0.02]
A-rated (w/o chair) = B-rated (w/o chair)	17.59*** [0.00]	11.11*** [0.00]	0.23 [0.63]	1.92 [0.17]

Figures in round parentheses denote standard errors.
Figures in square parentheses denote probability values under the null.
* Denotes significance at 10% levels.
** Denotes significance at 5% levels.
*** Denotes significance at 1% levels.

research chairs in our sample, and absolute levels of output (Total publications), below the A- and B-rated scholars without research chairs in our sample. The implication is thus that the funding intervention by the NRF was not targeted at the very strongest scholars in the South African academy (the A-rated scholars are consistently stronger in terms of absolute output, citations impact, and the joint *h*-index measure at the time of the award).

However, NRF chair holders, while not having the highest level and impact of research output, score higher than either A- or B-rated scholars in terms of citations per publication. It thus appears as if the award of the research chairs targeted scholars that have not yet reached maximum impact (levels of output and impact are not the highest), but which show promise (impact is higher than B-rated scholars, and impact per publication is higher even than A-rated scholars). Note that the findings are invariant to the use of ISI or Google Scholar based searches (contrast the ISI and PP columns of Table 3).

There is one important nuance to report in the results. When considering the NRF chairs' performance, we found that chairs who hold A- and B-ratings have demonstrably stronger performance compared to those with lower or no rating (who constitute more than 50% of the chairs). The NRF chairs with low or no rating had research performances worse than B-rated researchers without a chair.[22]

Since in the analytical work that follows this section, we also adjusted the ISI reported *h*-index by discipline-specific weights as suggested by Table II of Iglesias and Pecharromán (2007),[23] Table 3 reports the impact of the adjustment on the bibliometric measures in columns (4) and (6). While changing the absolute magnitude of the bibliometric measures, the inferences drawn regarding the relative rankings of the NRF chairs against rated scholars are unaffected.

[22] This finding is common across disciplines. The inference from the descriptive evidence is thus that the claim that the selection of researchers for the NRF chairs is based purely on narrow scholarly merit, needs qualification. This mirrors the earlier finding of Fedderke (2013) on 1932 researchers that the NRF rates. It is also the finding of Grimpe (2012) on the distribution of research grants across a sample of 800 researchers in Germany, when evaluated against the research output and impact of the researchers.

[23] The weights for the disciplinary categories in our study are as follows: biological 0.89; business 1.32; chemical 0.92; engineering 1.73; medical 0.67; physical 1.16; social 1.6.

4.1.2. Propensity score matching

To create control groups of researchers against which the NRF research chair holder's performance can be compared under the propensity score matching methodology, we begin with a pair of logit specifications that employ the performance of researchers in terms of the number of publications and total citations up to the point of the award of the research chairs in 2009, to determine the probability of a NRF chair being awarded given the objective scholarly performance of the researchers. In the second logit we also control for the principal disciplinary classification of researchers.[24]

The resultant implied probability values of each researcher are employed in order to classify researchers as having a Low, Medium, or High propensity of receiving a NRF research chair, irrespective of whether they in fact did so.

Mean values across our sample, for the standing of scholars at the time the NRF research chairs were awarded in 2009, are reported in Table 4. The mean values from the propensity scores suggest two principal empirical regularities. The first is that in terms of all the objective bibliometric performance measures, the probability of obtaining a research chair *declines* in rising objective research performance. Thus the highest number of publications, citations, average citations, and *h*-index is recorded for the group of researchers that has the lowest probability of obtaining a research chair, and the highest probability of obtaining a chair occurs for the group of researchers with the lowest number of publications and citations.

The second regularity is that within each of the propensity groups, the NRF has chosen researchers with a performance below that of researchers within that propensity group who were not awarded a research chair. The one exception is the group of researchers with the highest propensity of obtaining a research chair, where the research chair holders have a performance above that of comparable researchers who were not awarded a chair, though their performance remains below that of researchers with both a medium and low probability of receiving a chair.

Thus the NRF has chosen researchers with poor research productivity for their research chairs, though amongst the class of researchers with the very worst research performance, who have the highest probability of obtaining a chair, they have chosen those that are least poor.

4.1.3. Research trajectory of researchers

While the officially stated policy of the NRF was to award research chairs to "world class" researchers, the previous discussion shows that this is not borne out by the objective performance of research chair holders as measured by publications, citation, or *h*-index measures at the point of the research chair award.

How might we account for this disparity? One possibility, which is suggested by the relatively strong performance of NRF chairs under the average citations measure, is that the recipients of the chairs are researchers with a strong upward trajectory in their research performance, while still falling below the absolute level of output and impact of their more established colleagues.[25]

Table 7
Changing researcher performance with propensity score matching.

Change in:	Publications 2009–2012		Citations 2010–2012		Avg. citations 2010–2012		*h*-Index 2010–2012	
	(1) P&C	(2) P&C&D	(3) P&C	(4) P&C&D	(5) P&C	(6) P&C&D	(7) P&C	(8) P&C&D
Propensity score								
Low Prob w/o NRF	24.08*** (2.04)	21.92*** (2.08)	43.26*** (14.32)	37.51*** (14.04)	2.51*** (0.36)	2.33*** (0.35)	0.04 (0.08)	0.07 (0.08)
Med Prob w/o NRF	12.07*** (2.11)	10.68*** (2.13)	38.92*** (14.58)	33.94** (14.48)	2.18*** (0.39)	1.43*** (0.38)	0.13 (0.09)	0.19** (0.09)
High Prob w/o NRF	6.53*** (2.27)	10.64*** (2.38)	18.30 (15.62)	30.01* (15.98)	1.11*** (0.40)	2.12*** (0.43)	0.11 (0.09)	-0.01 (0.10)
NRF chair (All)	14.06*** (2.22)	12.89*** (2.29)	77.10*** (15.02)	72.10** (15.05)	2.58*** (0.39)	2.57*** (0.39)	0.20*** (0.09)	0.20** (0.09)
Non-NRF outlier	65.00*** (4.10)	66.65*** (4.18)	539.33*** (28.50)	544.13*** (28.18)	9.10*** (0.74)	9.45*** (0.74)	1.95*** (0.17)	1.95*** (0.17)
NRF outlier	79.66*** (7.22)	79.82*** (7.42)	700.61*** (49.48)	700.61*** (49.60)	16.76*** (1.29)	16.76*** (1.29)	3.94*** (0.30)	3.94*** (0.30)
Adj-R^2	0.59	0.57	0.66	0.66	0.54	0.53	0.51	0.51
N	300	300	300	300	300	300	300	300
χ^2 test of parameter equality								
Low Prob. w/o NRF = NRF chair (All)	11.56*** [0.00]	7.05*** [0.01]	2.66 [0.10]	3.70* [0.05]	0.02 [0.90]	0.21 [0.65]	1.76 [0.19]	1.12 [0.29]
Med Prob. w/o NRF = NRF chair (All)	0.43 [0.51]	1.07 [0.30]	3.33* [0.07]	4.27** [0.04]	0.50 [0.48]	4.34** [0.04]	0.36 [0.55]	0.01 [0.91]
High Prob. w/o NRF = NRF chair (All)	5.62*** [0.02]	0.98 [0.32]	7.37*** [0.01]	4.60** [0.03]	6.79*** [0.01]	0.61 [0.43]	0.48 [0.49]	2.57 [0.11]

Figures in round parentheses denote standard errors.
Figures in square parentheses denote probability values under the null.
P&C denotes Publications and Citations; P&C&D denotes Publications and Citations and Disciplines.
* Denotes significance at 10% levels.
** Denotes significance at 5% levels.
*** Denotes significance at 1% levels.

[24] Estimation results are reported in the online appendix.

[25] We also considered the possibility that the bias is due to a selection of younger researchers, and a reflection of inexperience in selection by universities with poor research cultures. However, the preponderance of all categories of research chair, including those with poor research track records, were selected by the two top-ranked research institutions in South Africa. While research chairs are marginally younger than the researchers in the control groups, the difference is small (consistently considerably less than a decade, often only a year younger). In addition, age differences show weak correlation with research performance differentials. An analysis of the researcher age and selecting university impact is available in the online appendix.

Table 8
Change in researcher performance from 2009 to 2012 with NRF chair breakdown by peer review rating and isolating citations to new work only.

Change in:	Publications 2009–2012 (1)	Citations 2010–2012 (2)	Avg. citations 2010–2012 (3)	*h*-Index 2010–2012 (4)
A-rated (w/o chair)	31.24*** (3.35)	159.30*** (25.45)	3.13*** (0.56)	0.43*** (0.13)
B-rated (w/o chair)	14.34*** (2.19)	57.07*** (16.63)	2.79*** (0.37)	0.22** (0.09)
NRF: A-rated	42.33*** (9.15)	148.22** (69.44)	2.41 (1.53)	0.11 (0.36)
NRF: B-rated	31.57*** (5.19)	222.07*** (39.37)	7.09*** (0.87)	0.43** (0.20)
NRF: Other rated	20.92*** (4.40)	82.36** (33.36)	2.38*** (0.74)	0.77*** (0.17)
Adj-R^2	0.08	0.06	0.06	0.02
N	300	300	300	300
	χ^2 test of parameter equality			
A-rated (w/o chair) = NRF: A-rated	1.30 [0.25]	0.02 [0.88]	0.19 [0.66]	0.72 [0.40]
A-rated (w/o chair) = NRF: Other rated	3.48* [0.06]	3.36* [0.07]	0.66 [0.42]	2.45 [0.12]
B-rated (w/o chair) = NRF: B-rated	9.37*** [0.00]	14.91*** [0.00]	20.72*** [0.00]	0.94 [0.33]
B-rated (w/o chair) = NRF: Other rated	1.80 [0.18]	0.46 [0.50]	0.26 [0.61]	8.38*** [0.00]
NRF: A-rated = NRF: B-rated	1.05 [0.31]	0.86 [0.35]	7.05*** [0.01]	0.60 [0.44]
NRF: A-rated = NRF: Other rated	4.45** [0.03]	0.73 [0.39]	0.00 [0.99]	2.78* [0.10]
NRF: B-rated = NRF: Other rated	2.45 [0.12]	7.33*** [0.01]	17.10*** [0.00]	1.67 [0.20]
A-rated (w/o chair) = B-rated (w/o chair)	17.80*** [0.00]	11.31*** [0.00]	0.25 [0.62]	1.93 [0.16]

Figures in round parentheses denote standard errors.
Figures in square parentheses denote probability values under the null.
* Denotes significance at 10% levels.
** Denotes significance at 5% levels.
*** Denotes significance at 1% levels.

To explore this possibility, in Table 5 we report the results of regressing the citations performance measure of researchers in our sample on a time trend over the 2005–2009 period under both the peer review and propensity score approaches. Relative to the control groups established under the NRF peer review mechanism, NRF research chairs had a stronger upward citations trajectory than B-rated researchers without NRF chairs, but weaker than that of A-rated researchers without NRF chairs. However, the upward citations trajectory is driven entirely by the NRF research chairs that are B- and A-rated. NRF research chairs that have ratings other than A or B (more than 50% of the chairs), have the lowest upward trajectory in our sample. It is therefore difficult to suggest that the NRF research chairs were chosen on the basis of a stronger upward research trajectory relative to the control group given by the NRF peer review mechanism.

Under the propensity score matching approach, there is some evidence that the research chairs were on strong upward trajectories at the point of selection, relative to the relevant control groups, since both low and high propensity NRF research chair holders show stronger positive trajectories than the corresponding controls without chairs. But the evidence is not unambiguous either. Note that the strongest positive trend occurs for chairs chosen from the low propensity grouping of researchers, while the positive trend for the chairs chosen from the high propensity grouping does not differ statistically from researchers in the low propensity grouping who did not receive a chair, and in the case of the medium propensity grouping the NRF chairs demonstrate no significant upward trend at all.[26]

To a significant extent, therefore, the downward bias in NRF research chair research productivity and impact remains unaccounted for. Nonetheless, to the extent that this evidence does establish the existence of strong positive research development at the point of the award on the part of chair recipients, note that the implication would be to strengthen the expectation that the research chair recipients should be able to significantly leverage off the significant additional funding of the chairs.

4.2. The performance change of NRF research chairs over the 2009–2012 period relative to the control groups

Our concern here is whether NRF research chairs holders report an appreciable increase in research output and impact, relative to researchers of comparable standing who did not receive a research chair. The focus is thus on H1.

4.2.1. NRF chair performance against the NRF peer review system control groups

We examine the *change* in the bibliometric measures for the NRF chairs, and A-rated and B-rated researchers without research chairs over the 2009–2012 period, by estimating specification (1) as reported in Table 6. For our change in research performance measure, we strictly eliminated all possible performance records that refer to work completed prior to 2009. Thus the change in performance measure is strictly attributable to the published work of the 2009–2012 period. Thus the data includes citations received in 2010–2012 for papers written from 2009 to 2012, number of publications from 2009 to 2012, average citations for the publications mentioned above and the change in the *h*-index. It is important to point out that the *h*-index is a cumulative measure associated to the researcher's prior publications, and cannot eliminate reference to work prior to 2009.

As the results of Table 6 show, all researchers, be they A-rated, B-rated, or NRF chair holders increased the number of recorded publications and total citations over the 2009–2012 period (see columns 1 and 2 of Table 6). The strongest increases in publications and in total citations are reported by A-rated researchers without NRF chairs. By contrast, B-rated researchers without NRF chairs, while also increasing their publications and citations counts, do so by a considerably lower margin than do A-rated researchers. NRF chair holders increased their publication count almost as much as A-rated researchers without NRF chairs (by 27.38 publications vs. 31.24), and total citations by a considerably closer margin

[26] One option would have been to include the trend structure in performance of researchers in the propensity score. Unfortunately, the number of time data points generated under the search engines per researcher is so limited as to render individual time trends subject to very substantial measurement error, especially given the time discontinuities of research output. This explains the use of the panel data evidence presented here.

than B-rated researchers (141.63 citations vs. 159.30 by A-rated researchers). The net result is that while the increase of B-rated researchers' publications was more than 50% lower than that of A-rated researchers, that of NRF chairs was approximately 10% lower. Statistically, it is the B-rated researchers without NRF chairs that differ from the two other groupings of researchers.

The average citations per publication increases for all categories of researcher with the NRF chairs showing the biggest increase (though not statistically significantly relative to A-rated researchers). A symmetrical pattern holds for the *h*-index results. See column 3 of Table 6.

The pattern in improved output and level of citations is mirrored in the recorded *h*-index. The strongest increase after 2009 occurs for NRF chairs (0.56), followed by A-rated researchers (0.43), while B-rated researchers show the smallest increase (0.22). See column 4 of Table 6.

Two sets of robustness checks were conducted on these results. First, we reestimated using a dependent variable (change in performance measure) which includes reference to work done prior to the award of the research chair. Second, we also employed discipline-normalized *h*-index values to assess the change in research performance. Under both robustness checks our findings are strictly symmetrical to those we have reported (under the discipline-normalized data the findings are in fact enhanced), and we therefore omit them for the sake of parsimony.[27]

What thus emerges is that the performance of NRF Chair awardees certainly changed after the granting of the chair: publications, citations, average citations per publication and the *h*-index rise, though not necessarily by as much as A-rated scholars. The findings thus speak to the success of a significant infusion of research funding on research productivity.

4.2.2. *NRF chair performance against the propensity score control groups*

As an alternative to the NRF peer review system ranking of scholars, we also examine the change in the bibliometric measures for the NRF chairs, compared to control groups generated with the propensity score matching of Section 4.1.2. To do so, we estimate specification (2), reported in Table 7. Reported results have eliminated all possible performance records that refer to work completed prior to 2009.

Recall that we compute propensity scores either using only publications and citations (P&C), or by using all of publications, citations and the disciplinary home of researchers (P&C&D) and we report results accordingly.

Consideration of the residuals in estimation indicated the presence of some strong outliers in the data. To test for the sensitivity of our results to the presence of these outliers, we also include dummy variables for the 10% top performers in both NRF and non-NRF researcher categories.[28]

As the results of Table 7 show, all researchers, be they of low, medium, high propensity control groups or actual NRF chair holders increased the number of recorded publications and total citations over the 2009–2012 period. The strongest increases in publications are reported by the low propensity control group *without* NRF research chair funding, followed by NRF chairs, and the medium and high propensity control groups. For Citations, Average

[27] The one exception is the Average citations measure, which on the change in full average citations count (including pre-2009 publications and citations) has a systematic (and clearly computable) downward bias relative to the measure which incorporates only post 2009 data. The downward bias is borne out in estimation on our data.

[28] Removing the controls for outliers, does not alter the inferences that emerge from our results, though controlling for outliers does improves goodness of fit diagnostics appreciably.

Table 9
Changing researcher performance with propensity score matching.

Change in:	Publications 2009–2012		Citations 2010–2012		Avg. citations 2010–2012		*h*-Index 2010–2012	
Propensity score	P&C (1)	P&C&D (2)	P&C (3)	P&C&D (4)	P&C (5)	P&C&D (6)	P&C (7)	P&C&D (8)
Low Prob w/o NRF	24.27*** (2.04)	21.96*** (2.10)	43.26*** (14.33)	37.51*** (14.07)	2.51*** (0.36)	2.33*** (0.35)	0.04 (0.08)	0.07 (0.08)
Med Prob w/o NRF	12.12*** (2.10)	10.69*** (2.14)	38.92*** (14.59)	33.94** (14.50)	2.18*** (0.40)	1.43*** (0.38)	0.13 (0.09)	0.19** (0.09)
High Prob w/o NRF	6.55*** (2.27)	10.65*** (2.38)	18.30 (15.63)	30.00* (16.01)	1.11*** (0.41)	2.12*** (0.42)	0.11 (0.09)	-0.01 (0.10)
Low Prob with NRF	21.30*** (5.29)	15.95*** (5.23)	87.96*** (32.91)	72.20** (33.05)	2.83*** (0.84)	2.75*** (0.83)	0.13 (0.19)	0.13 (0.19)
Med Prob with NRF	14.07*** (3.73)	13.56*** (3.91)	97.21*** (27.04)	59.79** (26.12)	2.53*** (0.66)	1.18* (0.68)	0.53*** (0.15)	0.26* (0.16)
High Prob with NRF	12.06*** (3.03)	13.46*** (3.12)	57.61*** (21.53)	91.10*** (21.39)	2.49*** (0.57)	3.39*** (0.55)	0.01 (0.13)	0.19 (0.14)
Non-NRF outlier	63.58*** (4.20)	66.31*** (4.28)	539.33*** (28.52)	544.13*** (28.24)	9.10*** (0.74)	9.45*** (0.73)	1.95*** (0.17)	1.95*** (0.17)
NRF outlier	75.80*** (7.64)	79.17*** (7.61)	705.78*** (49.93)	692.01*** (50.82)	16.83*** (1.31)	16.58*** (1.30)	3.90*** (0.30)	3.95*** (0.32)
Adj-R^2	0.59	0.57	0.66	0.66	0.53	0.54	0.52	0.51
N	300	300	300	300	300	300	300	300
χ^2 test of parameter equality								
Low Prob w/o NRF = Low Prob with NRF	0.29 [0.59]	1.19 [0.28]	1.55 [0.21]	0.93 [0.33]	0.12 [0.72]	0.21 [0.65]	0.20 [0.66]	0.08 [0.78]
Med Prob w/o NRF = Med Prob with NRF	0.21 [0.65]	0.41 [0.52]	4.04** [0.04]	0.75 [0.39]	0.20 [0.65]	0.11 [0.74]	5.19 [0.02]	0.16 [0.69]
High Prob w/o NRF = High Prob with NRF	2.11 [0.15]	0.52 [0.47]	2.18 [0.14]	5.23** [0.02]	3.87** [0.05]	3.32* [0.07]	0.41 [0.52]	1.49 [0.2]

Figures in round parentheses denote standard errors.
P&C denotes Publications and Citations; P&C&D denotes Publications and Citations and Disciplines.
Figures in square parentheses denote probability values under the null
* Denotes significance at 10% levels.
** Denotes significance at 5% levels.
*** Denotes significance at 1% levels.

Table 10
Change in researcher performance from 2009 to 2012 with NRF chair breakdown by discipline

Change in:	Publications 2009–2012 (1)	Citations 2010–2012 (2)	Avg. citations 2010–2012 (3)	*h*-Index 2010–2012 (4)
A-rated (w/o chair)	31.24*** (3.36)	159.30*** (25.42)	3.13*** (0.57)	0.43*** (0.13)
B-rated (w/o chair)	14.34*** (2.19)	57.07*** (16.61)	2.79*** (0.37)	0.22*** (0.08)
NRF chair in:				
Biology	21.22*** (6.01)	111.28** (45.50)	2.72*** (1.01)	0.30 (0.23)
Business &Economic	-1.17 (10.90)	-30.41 (82.53)	-0.34 (1.84)	-0.32 (0.41)
Chemical	20.12*** (7.91)	-30.13 (59.84)	-2.30* (1.33)	0.12 (0.30)
Engineering	11.58 (7.36)	74.40 (55.71)	3.26*** (1.24)	0.14 (0.28)
Medical	25.25*** (7.70)	188.13** (58.24)	4.32*** (1.30)	1.55*** (0.29)
Physical	16.74** (7.09)	188.75*** (53.64)	6.23*** (1.20)	0.34 (0.27)
Social	4.79 (6.95)	-3.21 (52.61)	0.81 (1.17)	0.17 (0.26)
Adj-R^2	0.06	0.04	0.01	0.01
N	300	300	300	300

Figures in round parentheses denote standard errors.
* Denotes significance at 10% levels.
** Denotes significance at 5% levels.
*** Denotes significance at 1% levels.

citations and the *h*-index results, the ordering is uniformly the same: the strongest increase is reported by NRF chair holders, followed by either the low or medium propensity control groups, with the high propensity control group without chairs uniformly reporting the smallest increase in research productivity.

Again the inference is thus that recipients of the NRF research chair funding did demonstrate improved performance after the funding grant. Relative to peers identified under the propensity score mechanism, the NRF chairs show greater improvement relative to the propensity score control groups (with the exception of the Publications measure).

However, note the implication of the tests for the equality of the parameters reported in Table 7. NRF chair performance is consistently statistically significantly better only with respect to the high propensity without research chair control group, which we have noted to have the very worst research performance in our sample. With respect to medium and low propensity without research chair control groups, statistically the NRF chairs do not show consistently higher performance, and in the case of publications, they statistically significantly underperform with respect to the low propensity without research chairs control group.

This qualification to the positive impact of the research chair funding is further confirmed by the fact that the ranking of the control groups again confirms that the selection of NRF chairs was perverse, in the sense that researchers least likely to have been chosen for the research chair funding consistently show strong performance improvements, even without the funding support under the chair mechanism.

4.3. Breaking down the NRF chair holders by scholarly standing

Across both the NRF peer review mechanism and the propensity score comparison, the implication is thus that the NRF research chairs have increased their research productivity after the award of their research funding – either by more than the strongest researchers in our sample (the low propensity of obtaining a chair control group under propensity score matching – though in terms of statistical significance this gain is insignificant), or almost as much (the A-rated researchers under the peer review mechanism).

In this section we turn to our H2 – examining to what extent the performance of NRF chairs is conditional on their scholarly standing at the point at which the funding is awarded.

4.3.1. NRF chair performance against the NRF peer review system control groups

As we have already noted, the NRF chair holders are not an homogeneous grouping. More than half of the chair holders have either no NRF rating of any kind, or a rating that falls below the most highly ranked A- or B-ratings.

A question of interest is therefore whether the increase in performance of NRF research chairs after the research grant is conditional on the level of scholarly experience and/or standing of the awardee. For this reason, we consider the results from the estimation of specification (3), which are reported in Table 8.[29]

The results confirm that the peer-reviewed standing of a researcher matters considerably in terms of the impact of research funding on performance. Specifically, the stronger the prior research standing of the researcher that receives the funding, the stronger the increase in research productivity after the award of the funding. While on average, across all NRF research holders, publications and citations after 2009 increased by less than A-rated researchers (see Table 6), this is not true for all types of holders of research chairs. Both A- and B-rated NRF research chair holders increased total publications by more than A-rated researchers without NRF chairs. In terms of increased total citations, the largest increase is reported by B-rated NRF research chair holders (222.07), though in this instance A-rated researchers without NRF chairs still outperform A-rated researchers with NRF chairs (159.30 vs. 148.22).[30]

The notable result here, however, is the substantial gap in performance between A- and B-rated NRF chair holders, and those chair holders who are rated other than A or B. These NRF chair holders do report an increase in publications, but only approximately 60% of the increase for A- and B-rated chair holders. For the increase in citations, the improvement lies between 37% (relative to B-rated chair holders), and 56% (relative to A-rated chair

[29] Results reported in Table 8 are those for which the change in performance measure strictly eliminates of all possible performance records that refer to work completed prior to 2009. The change in performance measure is thus strictly attributable to the 2009–2012 period. Again we conducted two robustness checks. First, we repeated our estimations using a change in performance measure that includes reference to work done prior to the award of the research chair. Second, we also estimated under discipline-specific normalizations of the performance measures. Results are strictly symmetrical under both robustness checks.

[30] Note that the downward bias in the average citations measure of performance is once again present, and for the same reason as previously.

holders). Thus the burden of success of the research chair initiative is falling disproportionately on less than 50% of the chair holders, who are principally responsible for any perceived increase in performance of the NRF research chairs. Statistically, it is the difference between NRF chairs with a B-rating and those with an Other rating that proves consistently significant.

Thus the averages for the NRF chair holders (in Section 4.2) are biased upward by the relatively strong performance of the A- and B-rated NRF chair holders, while the performance of unrated or lower rated chair holders is considerably weaker, even after the award of the chair. The prior scholarly standing of research chairs thus is of material importance to the subsequent performance of chair incumbents on average.

4.3.2. Breaking down the NRF chair holders by propensity score matching

Employing the propensity score matching methodology, we estimate specification (4) for our four measures of researcher performance, which we report in Table 9.

We again use both approaches to propensity score computation, and control for the presence of outliers.[31]

As the results of Table 9 show, all classes of researchers under the propensity score approach increased the number of recorded publications, citations and average citations over the review period.

Across all four research performance measures, NRF chair holders of any specified propensity score (low, medium, high), outperform the corresponding control group in terms of the raw coefficient values reported in Table 9. The only exception to this finding is that chair holders who had a low propensity of receiving a chair, do not always outperform their control group in publications (see columns 1 and 2).[32] Recall that the descriptive statistics showed that low propensities of receiving a research chair attached to the most productive researchers in our sample. The implication is thus that for the most productive researcher class, the additional funding also demonstrates the most questionable productivity gain over researchers without the funding.

Significantly, in terms of publications NRF chairs with a low or medium propensity of being chosen generally show stronger increases than those who had a high propensity of being chosen (recall that low and medium propensity researchers had higher research standing than high propensity researchers at the point of being chosen). Only in the case of citations and average citations do the high propensity research chairs outperform their low and medium propensity counterparts, and only if the disciplinary home of the researcher is included in the propensity computation.

The results of Table 9 suggest that NRF chairs did show performance improvements over their relevant control groups, to a larger extent than suggested by the peer review control groups of Section 4.3.1. But is the improved performance statistically significant? To test the significance of the difference between the NRF chairs and their control groups, we present tests for the equality of coefficients across the classes of researcher defined by the propensity to obtain NRF chairs (by χ^2 diagnostics). We report results in Table 9.

The diagnostics indicate that the only NRF chair grouping that does consistently statistically significantly better than their control, are the NRF chairs that had a high propensity of receiving an award, and only for citations and for average citations. Recall from the descriptive statistics of Section 4.1.2 that the high propensity group had the worst research performance in the sample, but that the chair recipients in this group were the least worst (i.e. had a better research performance than others in the high propensity group). Thus awarding funding to the best of the worst group, did allow them to outperform the other worst researchers. The only other significant exception to this finding is that under citations, medium propensity research chairs do better than their controls.

The use of propensity score matching thus strengthens the argument that even though the performance of NRF chairs improved after the granting of the chair, in most cases the change is not significantly different compared to scholars that did not receive significant funding and who constitute suitable controls. The exception is the case of the NRF chairs in the high probability group who outperform their peers in the sample in terms of citations and average citations and in some cases *h*-index.

The initially positive finding that NRF research holders do appear to substantially increase their research productivity, as reported in section 4.2, thus comes to be substantially qualified once NRF research chair performance is conditioned on the research standing of the chair awardee at the time of the award.

4.4. Breaking down the NRF chair holder performance by discipline

As a final step we also consider whether the distribution of research chairs across different disciplines is associated with differential research performance of the chair holders. To explore this question we estimate the specifications given by (5) including the NRF peer review controls, as reported in Table 10, and (6) including the propensity score controls, as reported in Table 11. This addresses H3 of the paper.

The results imply substantial disciplinary differences in the performance of the NRF chairs, confirming our third hypothesis. Strong increases in publications are concentrated in the Medical, Biological, Chemical and Physical sciences (in that order), while there is essentially small or even negative increase in the publications of NRF chairs in the Business and Economic, Engineering and Social sciences (on average). There is no difference in the results generated under the NRF peer review and the propensity score results in this inference.

With respect to citations, the performance of the disciplines is even more skewed. Increases are concentrated in the Medical and Physical sciences. While the Biological and Engineering sciences report positive (and strong) coefficients, only the Biological sciences are statistically significant, while the improvement in the Business and Economic, Chemical and Social sciences present a negative change. With respect to the *h*-index, it is only the Medical sciences that report a statistically significant, or for that matter meaningfully large performance increase.

Improved performance of NRF chairs thus appears to be concentrated in the Medical, Physical and Biological sciences.[33]

It is worth noting that the distribution of disciplinary performance shows a correspondence with how closely the NRF chairs showed strong performance at the time at which the chair was awarded. NRF chairs in the Medical, Physical and Biological sciences reported strong research track records in terms of bibliometric measures in 2009, at the time the chairs were granted. By contrast, for the Social, Engineering and particularly the Business and Economic sciences this correspondence was much weaker. The award of the NRF chairs does not appear to have changed this pattern.

[31] Again, results are not sensitive to the outlier controls.

[32] Where we fail to control for outliers, this extends to average citations and in terms of improved *h*-index scores.

[33] Use of the discipline-normalized Average citations and *h*-index values does not change any of the qualitative inferences drawn on the raw data.

Table 11
Change in researcher performance from 2009 to 2012 with NRF chair breakdown by discipline.

Change in:	Publications		Citations		Avg. citations		*h*-Index	
	(1) P&C	(2) P&C&D	(3) P&C	(4) P&C&D	(5) P&C	(6) P&C&D	(7) P&C	(8) P&C&D
Low Prob w/o NRF	32.49*** (2.90)	29.76*** (2.98)	132.10*** (22.74)	107.93*** (22.98)	3.26*** (0.50)	2.67*** (0.50)	0.22* (0.11)	0.19* (0.11)
Med Prob w/o NRF	14.67*** (3.10)	13.27*** (3.13)	82.07*** (24.21)	83.40*** (24.15)	3.76*** (0.53)	2.66*** (0.53)	0.39*** (0.12)	0.44*** (0.12)
High Prob w/o NRF	7.55** (3.35)	12.79*** (3.49)	35.16 (26.20)	65.11** (26.91)	1.40** (0.58)	3.49*** (0.59)	0.23* (0.13)	0.21 (0.13)
NRF chair in:								
Biology	21.22*** (5.86)	21.22*** (30.76)	111.28** (45.84)	111.28** (46.34)	2.72*** (1.00)	2.72*** (1.01)	0.30 (0.23)	0.30 (0.23)
Business &Economic	-1.17 (10.62)	-1.17 (10.89)	-30.41 (83.14)	-30.41 (84.04)	-0.34 (1.81)	-0.34 (1.83)	-0.32 (0.41)	-0.32 (0.41)
Chemical	20.12*** (7.70)	20.12*** (7.89)	-30.13 (60.28)	-30.13 (60.94)	-2.30* (1.31)	-2.30* (1.33)	0.12 (0.30)	0.12 (0.30)
Engineering	11.58 (7.17)	11.58 (7.35)	74.40 (56.13)	74.40 (56.74)	3.26*** (1.22)	3.26*** (1.24)	0.14 (0.28)	0.14 (0.28)
Medical	25.25*** (7.50)	25.25*** (7.69)	188.13*** (58.68)	188.13*** (59.31)	4.32*** (1.28)	4.32*** (1.30)	1.55*** (0.29)	1.55*** (0.29)
Physical	16.74** (6.91)	16.74** (7.08)	188.75*** (54.04)	188.75*** (54.63)	6.23*** (1.18)	6.23*** (1.12)	0.34 (0.27)	0.34 (0.27)
Social	4.79 (6.77)	4.79 (6.94)	-3.21 (53.00)	-3.21 (53.58)	0.81 (1.16)	0.81 (1.17)	0.17 (0.26)	0.17 (0.26)
Adj-R^2	0.11	0.07	0.03	0.01	0.03	0.01	0.01	0.01
N	300	300	300	300	300	300	300	300

Figures in round parentheses denote standard errors.
* Denotes significance at 10% levels.
** Denotes significance at 5% levels.
*** Denotes significance at 1% levels.

4.5. *Adding the cost dimension*

We have seen that the granting of the NRF chairs did have a statistically significant impact on the bibliometric measures of associated researchers.

But recall that at the time of the research chair program creation A-rated researchers received NRF grants valued at US$10,000 per annum, B-rated researchers grants of US$8000, while the budgetary allocation for NRF chairs is US$300,000 per annum (approximately: we use an exchange rate of ZAR10:US$1).[34] In Table 12 we report the implied costs per additional publication and citation after 2009 across the various researcher rating categories. The very substantial differential in budgetary allocation across the various classes of researcher, and the limited increased productivity of the NRF research chairs over other researchers, results in a strong cost differential per publication and per citation between NRF research chairs and A- or B-rated researchers. This is irrespective of whether the NRF chair is A-, B- or otherwise rated. The strong cost differential per publication and per citation between NRF chairs and their peers is also maintained when using propensity scores.

The inference is that while the expenditure on the NRF chairs has increased output, it has come at some cost – and the cost is substantial. As Table 12 demonstrates, the smallest differential in cost between the NRF chairs and the A- and B-rated researchers, is a factor of 13:1 per publication, and 10:1 per citation. And the differential rises to 45:1 in the case of Publications, and 58:1 in the case of citations. When using propensity scores including disciplines, for which we utilize an average of $9,000 for their attributed research funding, the smallest differential in cost between the NRF chairs and the non-NRF chairs in each control group, is a factor of 17:1 per publication, and 12:1 per citation. And the differential rises to 26:1 in the case of Publications, and 47:1 in the case of citations.

In short, the implication is that if the objective is to raise the research output of the academy-based research system, a more productive means of doing so may well be by allocating more funding at the margin to A-rated and B-rated researchers, rather than allocating the very substantial budgetary resources associated with the NRF research chairs.

5. Discussion and policy inferences

The evidence from our data suggests that substantial research funding is associated with higher numbers of publications and citations, though the increase is moderate when compared to the performance of researchers of equivalent standing. Given the moderate productivity gain, factoring in the cost of the productivity gain realized by NRF research chairs, shows that the cost per unit of additional research output for the chairs proves to be substantial.

Our results also demonstrate that the performance changes associated with NRF research chair funding is strongly conditional on the scholarly standing of the recipient. Scholars that are highly ranked in terms of peer review or that have high levels of underlying performance demonstrate a far higher rate of return on investment in terms of research output than those that have a low peer review ranking.

In addition, strong funding increases do not show the same rate of return across all disciplines. For NRF chair holders Publications increase statistically significantly only in the Biological, Chemical, Medical and Physical sciences, and citations only in the Biological, Medical and Physical sciences. Chair holders in the Business and

[34] We do not have data on other sources of funding available to researchers. However, given the status that attaches to the NRF research chairs, there is no reason to suppose that holders of chairs have a lower capacity to raise research funding than other researchers.

Table 12
Cost of additional productivity across NRF researcher rating categories in US dollars and ratio of costs per publication and citation between NRF research chairs and A and B-rated researchers

Peer review comparison			Propensity score comparison (on P&C&D selection)		
	Per new publication	Per new citation		Per new publication	Per new citation
A-rated w/o chair	320	63	Low (w/o NRF)	302	83
B-rated w/o chair	558	140	Med (w/o NRF)	678	108
A-rated NRF chair	7087	2024	High (w/o NRF)	704	138
B-rated NRF chair	9503	1351	Low (with NRF)	6064	1824
Other Rated NRF chair	14,340	3643	Med (with NRF)	17,647	5018
			High (with NRF)	12,024	1647
Ratio					
A-rated NRF chair: A-rated w/o chair	22: 1	32: 1	Low with NRF: Low w/o NRF:	20: 1	22: 1
A-rated NRF chair: B-rated w/o chair	13: 1	14: 1	Med with NRF: Med w/o NRF:	26: 1	47: 1
B-rated NRF chair: A-rated w/o chair	30: 1	22: 1	High with NRF: High w/o NRF:	17: 1	12: 1
B-rated NRF chair: B-rated w/o chair	17: 1	10: 1			
Other Rated NRF chair: A-rated w/o chair	45: 1	58: 1			
Other Rated NRF chair: B-rated w/o chair	26: 1	26: 1			

Economic, Engineering or the Social sciences show no statistically discernible increase in either dimension.

We draw a number of policy conclusions from this evidence.

Funding allocation should be more responsive to past research performance: What our evidence shows is that while funding raises research output, the impact of funding is enhanced where it is allocated under selectivity at the point of award. The impact of the research chair funding is at its highest when allocated to researchers with A- or B-ratings, or those that have the highest productivity record under the propensity score approach. By contrast, NRF chair holders with lower ratings perform worse than high productivity researchers who did not receive the research chair funding. Thus choosing researchers with strong performance records at the point at which the award is made, generates a stronger research productivity increase than choosing researchers with poor research track records. The implication is that reallocating research funding from low productivity to high productivity researchers will generate gains in research output.

This policy inference is currently not obviously followed by public research funding agencies. We reported that for the South African funding agency the attention paid to selectivity is poor. The majority of NRF research chair recipients were of low research ratings. In this, the initiative has some similarities with reported biases in the Canada Chair initiative and in Germany.[35] The cost for these funding agencies is foregone research output.

Funding should be given more broadly: Consideration of the differential cost per unit of research output or impact between researchers who do, and those who do not receive research chair funding, implies that the marginal returns from increased funding are savagely diminishing. Even for the most productive NRF chair recipients, the cost per additional publication is 22 times as high as comparable researchers without the funding, and per citation 32 times as high (compare A-rated researchers with and without chairs). Under such strongly falling marginal productivity, the considerably higher funding of chairs simply does not generate a commensurate research productivity response. It follows that if the objective of research funding is to raise the level of research output and impact, heavy focus on a small group of researchers is unlikely the most effective funding model to adopt. Spreading funding more equally across all highly rated researchers may elicit a stronger output response than bunching resourcing on a few individuals to the exclusion of all others (in our data, there were a considerable number of researchers without chairs). Once the funding is guaranteed over substantial periods of time, negative incentive effects are only likely to compound this problem.

Selectivity should continue beyond the initial award: There is no reason why selectivity should be restricted to the point of award - in at least two senses. First, if the objective is to maximize research output from the allocation of necessarily limited and scarce resources, monitoring ongoing performance of researchers after the award, evaluating performance, and making funding conditional on performance significantly above that of comparable peer groups who did not receive the funding support, would serve as an incentive to maximize the return from the investment from the public funding agency. If a small group of researchers are to be singled out through disproportionate funding support, it is appropriate that the level of support be justified in terms of exceptional productivity levels. Continued funding should be strictly tied to exceptional performance, and poor performance should be strictly tied to the withdrawal of funding.

Second, selectivity after the award of initial funding, can be used as a mechanism to scale funding upwards gradually, if and only if the productivity performance of the researcher justifies the increased funding, and again conditional on monitoring that the increased funding is associated with continued productivity gains that merit the higher funding.

Accountability and transparency of the criteria on which grants are awarded should be improved: The award of the substantial research funding to the NRF chairs appears to be biased towards the least productive researchers in our sample, despite the publicly stated goal to the contrary. Moreover, this bias appears to be at best only weakly associated with strong upward trajectories in prior research performance, which renders the bias difficult to account for. If the selection of the research chairs *was* on the basis of strong positive trajectories, the expectation is of an enhanced positive impact of the funding, making the weakness of this effect all the more surprising. Such lack of transparency in the allocation of public funding, and a bias that appears to run against the stated goals of the program compromises the efficiency of the program in achieving its ends, and can come to damage the legitimacy of the funding agency in the eyes of researchers. Particularly since similar concerns have been voiced in public funding programs elsewhere (Canada, Germany), a general policy recommendation that follows is that public funding agencies in research funding improve

[35] The effectiveness of the Canadian research chair initiative has been called into question. The program has been criticized for being influenced by political imperatives and the preferences and choices of university administrators. In addition, an analysis of publications and citations of Canadian research chairs during the first five years of the program showed that the group of scholars who received Canadian research chairs is heterogeneous, with some obvious high performers but also with many with publishing and citation records similar or inferior to comparable non research chair holders, Siler and Mclaughlin (2008). On Germany see Grimpe (2012).

accountability by increasing the transparency of the criteria on which grants are awarded. Publication of the objective research performance of grant recipients at the point of award, relative to that of applicants whose applications proved unsuccessful would be one means of improving transparency, without having to compromise the ability of the agency to select innovative proposals from new researchers without prior research track records.

Allocate funding differentially across disciplines: Since the impact of big funding for research has a differential success across disciplines, policy might also fruitfully consider stricter allocative criteria across disciplines with proven rates of return on investment. However, provided that selectivity in the funding process is implemented per individual researcher, the disciplinary selectivity would be of secondary importance, since only successful research initiatives would receive funding.

Overall, conditional on the objective of public research funding agencies really being the promotion of research output of high quality and impact, the findings of this paper carry implications for the funding models that such agencies adopt. In contrast to the research chair approach, which concentrates research funding ex ante, employ broad-based funding for as large a group of researchers who meet minimum quality standards at the outset of a funding initiative; then increase funding to those researchers who demonstrate objective performance improvements in terms of both quantity and quality of research after the initial funding awards; increasingly concentrate funding on the most promising research initiatives over time. As a result, both adverse selection and moral hazard are reduced, the most productive research initiatives are rewarded, and any disciplinary differentials in terms of the productivity impact of research funding (such as those we report) can be fully accommodated, without having to make a priori prioritization decisions across disciplines.

Funding allocation should thus follow revealed productivity. A corollary is that for the associated incentive mechanisms to work, productivity has to be transparently and objectively monitored. The reluctance of public research funding bodies to embrace the use of the growing number of objective bibliometric measures, whose collection is greatly facilitated by the growth in information technology, serves to constrain the effectiveness of research funding. Use of such objective measures at least in part in reaching allocative decisions, is an obvious remedy.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.respol.2014.09.009.

References

Abramo, G., D'Angelo, C.A., 2007. Measuring science: irresistible temptations, easy shortcuts and dangerous consequences. Current Science 93 (6), 762–766.

Abramo, G., D'Angelo, C.A., Caprasecca, A., 2009. Allocative efficiency in public research funding: can bibliometrics help? Research Policy 38, 206–215.

Aghion, P., Fedderke, J.W., Howitt, P., Viegi, N., 2013a. Testing creative destruction in an opening economy: the case of the South African manufacturing industries. Economics of Transition 21 (3), 419–450.

Aghion, P., Van Reenen, J., Zingales, L., 2013b. Innovation and institutional ownership. American Economic Review 103 (1), 277–304.

Aksnes, D.W., Taxt, R.E., 2004. Peer reviews and bibliometric indicators: a comparative study at Norwegian Universities. Research Evaluation 13 (1), 33–41.

Archambault, E., Gagné, E.V., 2004. The Use of Bibliometrics in Social Sciences and Humanities. Social Sciences and Humanities Research Council of Canada (SSHRCC), Montreal.

Belew, R.K., 2005. Scientific impact quantity and quality: analysis of two sources of bibliographic data, v1., pp. 11, arXiv:cs.IR/0504036.

Bornmann, L., 2011. Scientific peer review. Annual Review of Information Science and Technology 45, 199–245.

Bornmann, L., Daniel, H.-D., 2005. Does the h-index for ranking of scientists really work? Scientometrics 65 (3), 391–392.

Bornmann, L., Daniel, H.-D., 2007. What do we know about the h index? Journal of the American Society for Information Science and Technology 58 (9), 1381–1385.

Bornmann, L., Marx, W., Schier, H., Rahm, E., Thor, A., Daniel, H.D., 2009. Convergent validity of bibliometric Google Scholar data in the field of chemistry. Citation counts for papers that were accepted by Angewandte Chemie International Edition or rejected but published elsewhere, using Google Scholar, Science Citation Index, Scopus, and Chemical Abstracts. Journal of Informetrics 3 (1), 27–35.

Bosman, J., Mourik, I., van Rasch, M., Sieverts, E., Verhoeff, H., 2006. Scopus Reviewed and Compared. The Coverage and Functionality of the Citation Database Scopus, Including Comparisons with Web of Science and Google Scholar. Utrecht University Library, Utrecht, http://igitur-archive.library.uu.nl/DARLIN/2006-1220-200432/Scopus doorgelicht&vergeleken-translated.pdf.

Braun, D., 1998. The role of funding agencies in the cognitive development of science. Research Policy 27, 807–821.

Butler, L., 2003. Explaining Australia's increased share of ISI publications – the effects of a funding formula based on publication counts. Research Policy 32, 143–155.

Butler, L., 2006. RQF Pilot Study Project – History and Political Science Methodology for Citation Analysis. http://www.chass.org.au/papers/bibliometrics/CHASS_Methodology.pdf

Cassiman, B., Veugelers, R., 2006. In search of complementarity in innovation strategy: internal R&D and external knowledge acquisition. Management Science 52 (1), 68–82.

Crépon, B., Duguet, E., Mairesse, J., 1998. Research investment innovation and productivity: an econometric analysis. Economics of Innovation and New Technology 7 (2), 115–158.

Cronin, B., Meho, L., 2006. Using the h-index to rank influential information scientists. Journal of the American Association for Information Science and Technology 57 (9), 1275–1278.

Derrick, G.E., Sturk, H., Haynes, A.S., Chapman, S., Hall, W.D., 2010. A cautionary bibliometric tale of two cities. Scientometrics 84 (2), 317–320.

Dixit, A.K., Pindyck, R.S., 1994. Investment Under Uncertainty. University Press, Princeton.

Ederer, F., Manso, G., 2012. Is Pay-for-Performance Detrimental to Innovation? Anderson School of Management: UCLA, Mimeo.

Egghe, L., Rousseau, R., 2006. An informetric model for the Hirsch-index. Scientometrics 69 (1), 121–129.

Fagerberg, J., 1994. Technology and international differences in growth rates. Journal of Economic Literature 32 (3), 1147–1175.

Fedderke, J.W., 2013. The Objectivity of National Research Foundation peer review in South Africa assessed against bibliometric indexes. Scientometrics, http://dx.doi.org/10.1007/s11192-013-0981-0.

Falagas, M.E., Pitsouni, E.I., Malietzis, G.A., Pappas, G., 2008. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. The FASEB Journal 22, 338–342, http://dx.doi.org/10.1096/fj.07-9492LSF.

García-Pérez, M.A., 2010. Accuracy and completeness of publication and citation records in the Web of Science, PsycINFO, and Google Scholar: a case study for the computation of h indices in psychology. Journal of the American Society for Information Science and Technology 61 (10), 2070–2085.

Garriga, H., Von Krogh, G., Spaeth, S., 2013. How constraints and knowledge impact on innovation. Strategic Management Journal 34, 1134–1144.

Glänzel, W., 2006. On the opportunities and limitations of the H-index. Science Focus 1 (1), 10–11.

Gray, J.E., Hamilton, M.C., Hauser, A., Janz, M.M., Peters, J.P., Taggart, F., 2012. Scholarish: Google Scholar and its value to the sciences. Issues in Science and Technology Librarianship 70 (Summer), http://dx.doi.org/10.5062/F4MK69T9.

Grimpe, C., 2012. Extramural research grants and scientists' funding strategies: beggars cannot be choosers? Research Policy 41, 1448–1460.

Harzing, A.-W., 2007–2008. Google Scholar – A New Data Source for Citation Analysis. http://www.harzing.com/pop_gs.htm

Harzing, A.-W., 2008. Reflections on the H-Index. http://www.harzing.com/pop_hindex.htm

Hicks, D., 2012. Performance-based university research funding systems. Research Policy 41, 251–261.

Hirsch, J.E., 2005. An index to quantify an individual's scientific research output, v5, arXiv:physics/0508025.

Horrobin, D.F., 1990. The philosophical basis of peer review and the suppression of innovation. Journal of the American Medical Association 263, 1438–1441.

Iglesias, J.E., Pecharromán, C., 2007. Scaling the h-index for different scientific ISI fields. Scientometrics 73 (3), 303–320.

Jacsó, P., 2005. Google Scholar: the pros and the cons. Online Information Review 29 (2), 208–214.

Jacsó, P., 2006a. Dubious hit counts and cuckoo's eggs. Online Information Review 30 (2), 188–193.

Jacsó, P., 2006b. Deflated, inflated and phantom citation counts. Online Information Review 30 (3), 297–309.

Jacsó, P., 2010. Metadata mega mess in Google Scholar. Online Information Review 34 (1), 175–191.

Klingebiel, R., Rammer, C., 2013. Resource allocation strategy for innovation portfolio management. Strategic Management Journal, http://dx.doi.org/10.1002/smj.

Korrevaar, J.C., Moed, H.F., 1996. Validation of bibliometric indicators in the field of mathematics. Scientometrics 36 (3), 343–362.

Kousha, K., Thelwall, M., 2007. Google Scholar Citations and Google Web/URL citations: a multi-discipline exploratory analysis. Journal of the American Society for Information Science and Technology 58 (7), 1055–1065.

Kousha, K., Thelwall, M., 2008. Sources of Google Scholar citations outside the Science Citation Index: a comparison between four science disciplines. Scientometrics 74 (2), 273–294.
Kulkarni, A.V., Aziz, B., Shams, I., Busse, J.W., 2009. Comparisons of citations in Web of Science, Scopus, and Google Scholar for articles published in general medical journals. Journal of the American Medical Association 302 (10), 1092–1096.
Laursen, K., Salter, A., 2006. Open for innovation: the role of openness in explaining innovation performance among UK manufacturing firms. Strategic Management Journal 27 (2), 131–150.
Leiblein, M.J., Madsen, T.L., 2009. Unbundling competitive heterogeneity: incentive structures and capability influences on technological innovation. Strategic Management Journal 30, 711–735.
Leiponen, A., Helfat, C., 2010. Innovation objectives knowledge sources and the benefits of breadth. Strategic Management Journal 31 (2), 224–236.
Leiponen, A., Helfat, C., 2011. Location, decentralization, and knowledge sources for innovation. Organization Science 22 (3), 641–658.
Li, H., Atuahene-Gima, K., 2001. Product innovation strategy and the performance of new technology ventures in China. Academy of Management Journal 44 (6), 1123–1134.
Mairesse, J., Mohnen, P., 2002. Accounting for innovation and measuring innovativeness: an illustrative framework and an application. American Economic Review 92 (2), 226–230.
Martin, B.R., 1996. The use of multiple indicators in the assessment of basic research. Scientometrics 36 (3), 343–362.
Meho, L.I., Yang, K., 2007. A New Era in citation and bibliometric analyses: Web of Science, Scopus, and Google Scholar. Journal of the American Society for Information Science and Technology 58 (13), 2105–2125.
Mokyr, J., 2002. The Gifts of Athena: Historical Origins of the Knowledge Economy. University Press, Princeton.
Mowery, D.C., Rosenberg, N., 1989. Technology and the Pursuit of Economic Growth. University Press, Cambridge.
Moxhan, H., Anderson, J., 1992. Peer review. A view from the inside. Science and Technology 36 (3), 293–310.
Nelson, R.R., 1992. What is "commercial" and what is "public" about technology, and what should be? In: Rosenberg, N., Landau, R., Mowery, D.C. (Eds.), Technology and the Wealth of Nations. University Press, Stanford, pp. 57–72.
Nelson, R.R., 1996. The Sources of Economic Growth Cambridge Mass. Harvard University Press, Cambridge.
Nelson, R.R., Wright, G., 1992. The rise and fall of American technological leadership: the Postwar Era in historical perspective. Journal of Economic Literature 30 (4), 1931–1964.
Neuman, S., 2003. Systemic discrimination and the Canada research chairs: diagnosis and treatment. Clinical and Investigative Medicine 26 (1), 35–37.
Nisonger, T.E., 2004. Citation autobiography: an investigation of ISI database coverage in determining author citedness. College & Research Libraries 65 (2), 152–163.
Oppenheim, C., 1997. The correlation between citation counts and the 1992 research assessment exercise ratings for British research in genetics, anatomy and archaeology. Journal of Documentation 53, 477–487.
Polster, C., 2002. A break from the past: impacts and implications of the Canada foundation for innovation and the Canada research chairs initiatives. The Canadian Review of Sociology and Anthropology 39 (3), 275–299.
Rehn, C., Kronman, U., Wadskog, D., 2007. Bibliometric Indicators – Definitions and Usage at Karolinska Institutet. Karolinska Institutet University Library, Stockholm, Sweden.
Rinia, E.J., Van Leeuwen, Th.N., Van Vuuren, H.G., Van Raan, A.F.J., 1998. Comparative analysis of a set of bibliometric indicators and central peer review criteria: evaluation of condensed matter physics in the Netherlands. Research Policy 27, 95–107.
Roediger III, H.L., 2006. The h index in science: a new measure of scholarly contribution. APS Observer: The Academic Observer 19 (4) http://www.psychologicalscience.org/index.php/publications/observer/2006/april-06/the-h-index-in-science-a-new-measure-of-scholarly-contribution.html
Rosenberg, N., 1994. Exploring the Black Box: Technology, Economics and History. University Press, Cambridge.
Siler, K., Mclaughlin, N., 2008. The Canada research chairs program and social science reward structures. Canadian Review of Sociology 45 (1), 93–119.
Testa, J., 2004. The Thomson Scientific Journal Selection Process. http://scientific.thomson.com/free/essays/selectionofmaterial/journalselection/
Van Raan, A.F.J., 2005. Fatal attraction: conceptual and methodological problems in the ranking of universities by bibliometric measures. Scientometrics 62 (1), 133–143.
Van Raan, A.F.J., 2006. Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgement for 147 chemistry research groups. Scientometrics 67 (3), 491–502.
Van Reenen, J., 2011. Does competition raise productivity through improving management quality? International Journal of Industrial Organization 29 (3), 306–316.
Vaughan, L., Shaw, D., 2008. A new look at evidence of scholarly citations in citation indexes and from web sources. Scientometrics 74 (2), 317–330.