
http://www.diva-portal.org

This is the published version of a paper published in *Journal of Informetrics*.

Citation for the original published paper (version of record):

Van Den Besselaar, P., Heyman, U., Sandström, U. (2017)
Perverse effects of output-based research funding? Butler's Australian case revisited.
Journal of Informetrics, 11(3): 905-918
https://doi.org/10.1016/j.joi.2017.05.016

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-214064

Journal of Informetrics xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Journal of Informetrics

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

Regular article

Perverse effects of output-based research funding? Butler's Australian case revisited

Peter van den Besselaar [a,*], Ulf Heyman [b], Ulf Sandström [c]

[a] Network Institute, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[b] Uppsala University, Uppsala, Sweden
[c] KTH Royal Institute of Technology, Stockholm, Sweden

ARTICLE INFO

Article history:
Accepted 2 November 2015
Available online xxx

ABSTRACT

More than ten years ago, Linda Butler (2003a) published a well-cited article claiming that the Australian science policy in the early 1990s made a mistake by introducing output based funding. According to Butler, the policy stimulated researchers to publish more but at the same time less good papers, resulting in lower total impact of Australian research compared to other countries. We redo and extend the analysis using longer time series, and show that Butlers' main conclusions are not correct. We conclude in this paper (i) that the currently available data reject Butler's claim that "journal publication productivity has increased significantly... but its impact has declined", and (ii) that it is hard to find such evidence also with a reconstruction of her data. On the contrary, after implementing evaluation systems and performance based funding, Australia not only improved its share of research output but also increased research quality, implying that total impact was greatly increased. Our findings show that if output based research funding has an effect on research quality, it is positive and not negative. This finding has implications for the discussions about research evaluation and about assumed perverse effects of incentives, as in those debates the Australian case plays a major role.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

More than ten years ago, Linda Butler (2003a) published a well-cited article analyzing the effects of the increased emphasis on research evaluation and of the introduction of output based funding in the Australian academic research system during the first half of the 1990s. The science policy during that time included university research funding that was partially based on the number of publications. According to Butler, this policy stimulated researchers to publish more but at the same time less good papers. To illustrate this, she showed that Australian number of papers was increasing, as was the share in world production of papers, but that the relative citation impact of those publications did not increase. As the same indicators for other countries were increasing, Butler concluded that the Australian knowledge production was losing quality. She also used the changing distribution of publications over quartiles of the journal impact factor (JIF) to show that the increase of the number of papers mainly occurred in low impact journals (Butler, 2002).

* Corresponding author.
E-mail addresses: p.a.a.vanden.besselaar@vu.nl (P. van den Besselaar), ulf.heyman@uadm.uu.se (U. Heyman), ulf.sandstrom@indek.kth.se (U. Sandström).

http://dx.doi.org/10.1016/j.joi.2017.05.016
1751-1577/© 2017 Elsevier Ltd. All rights reserved.

Please cite this article in press as: van den Besselaar, P., et al. Perverse effects of output-based research funding? Butler's Australian case revisited. *Journal of Informetrics* (2017), http://dx.doi.org/10.1016/j.joi.2017.05.016

Butler suggested two behavioral mechanisms to explain this finding: "Increased system-wide and institutional performance evaluation based on aggregate output measures appears to be altering researchers' publication habits" (Butler, 2003a, p. 154). Firstly, she claimed that the new policy was stimulating Australian researchers to select on average lower level journals for their increased output. "When this element (of output based funding) was incorporated into the funding formulae in 1995, universities and researchers were quick to calculate the 'value' of a publication", and with "no differentiation between the quality or impact of the publications, there is little incentive to strive for placement in a prestigious journal" (Butler, 2002, p. 877). Secondly, also because of the new output oriented policy, one would "expect 'publication inflation' from a performance- based system where aggregate publication counts are a key component" (Butler, 2003a, p. 154). The concept of publication inflation was introduced by researchers at the Science Policy Research Unit (SPRU) in Brighton and hinted at phenomena such as "salami publishing and game playing". According to Butler, academics accepted a situation with growing output, but an output appearing in lower impact journals (Butler, 2003a, p. 154): "In consequence, journal publication productivity has increased significantly in the last decade, but its impact has declined" (Butler, 2003a, p. 143), a conclusion that has been cited many times since. The policy lesson was also obvious, as "a more detailed examination of the data reveals that Australia's *RCI* continues to decline, and raises important questions on the wisdom of a policy that rewards quantity, with scant regard to quality" (Butler, 2003a, p. 143).

The question, however, is whether these observations and conclusions are in accordance with available data and present knowledge. There is not much direct evidence available for behavioral reactions on 'perverse incentives', or for 'salami slicing' practices, whereas research does find high commitment and motivation of researchers (Van der Weijden, Belder, van Arensbergen, & van den Besselaar, 2015) and a positive correlation between commitment, motivation and productivity (Pelz & Andrews, 1966). Other studies indicate that, on average, the more papers a researcher publishes, the higher the proportion of these papers that are amongst the most cited (Van den Besselaar & Sandström, 2015; Sandström & van den Besselaar, 2016). This holds more clearly for established researchers than for early career researchers (Larivière & Costas, 2015). The positive relation between number of papers and proportion of highly cited papers is in line with theories about scientific creativity (Simonton, 2004).

Butler's argument has been repeated many times (e.g. Geuna & Martin, 2003; Hicks, 2009; OECD, 2010; Stephan, 2012; Hicks et al., 2015), and has become common knowledge in science policy studies. For example, Schneider, Aagaard, and Bloch (2016) use the Australian case as a frame of reference for their study of the effects of the so-called 'Norwegian model'. Especially concerning government funding of universities, Butler's papers have been very influential. In view of the scarce behavioral evidence and the quite short time series available to Butler, it seems important to take a fresh look at the Australian policy intervention and the effects of it.

Doing the analysis today has several advantages viz. Butler's study, as citations counts will be rather stable, the database has been improved, better indicators have developed, and the time since changes in the funding system is long enough to be sure that possible changes in publication numbers and quality can be detected. Our aim is thus not to replicate Butler's study, but to reanalyze the effect of the changes in the funding system in Australia during the first half of the nineties.

We conclude in this paper (i) that the currently available data reject Butler's claim that "journal publication productivity has increased significantly in the last decade, but its impact has declined" (Butler, 2003a) and (ii) that it is hard to find evidence for this also with a reconstruction of her data. Indeed, our evidence suggests that the average impact per publication has increased after emphasis on evaluation became stronger and performance based funding was introduced. So Australia improved its relative share of research output without losing quality and thus the total impact was greatly increased.

It may be useful to clarify here the use of citation impact as proxy for scholarly quality, as this is disputed in the literature (e.g., Martin & Irvine, 1983, p. 67–71; MacRoberts & MacRoberts, 1989). It is important to keep in mind that this criticism mainly holds for the use of bibliometric indicators at the individual level (Van Raan, 1996). At the individual level, more dimensions of scholarly quality than citation impact play a role (Van Arensbergen, Van der Weijden, & van den Besselaar, 2014), such as *independence*, *originality,* and *creativity* – dimensions that have to a large extent been neglected by bibliometricians. Elsewhere we have taken up this challenge (Van den Besselaar, Sandström, & van der Weijden, 2012). Finally, also other quality dimensions are relevant for which indicators can be developed, such as societal impact (de Jong, Barker, Cox, Sveinsdottir, & Van den Besselaar, 2014; Van der Weijden, Verbree, & van den Besselaar, 2012). However, it is also well known that if one uses data on larger groups (teams, universities, or countries), citations are a fairly reliable and valid proxy for scholarly quality (e.g. Narin, 1976; Roche & Smith, 1978; Nederhof & van Raan, 1993; Phelan, 1999), and this is the way we use citation impact in this paper – as did Butler (2003a) .

2. The Australian incentive system in the 1990s

Since long, Australia has a funding system consisting of block grants and competitive research grants via research agencies. Core funding (block grants) became in 1990 dependent on a "Research Quantum" (RQ) based on success in acquiring grants, i.e. research earnings. This funding model later developed as student numbers and publication components were added to the formula. In 1995 it was for the first time announced that a new Composite Index (CI) would be introduced (Butler, 2002, 2003a, 2003b, 2004) which also included publication counts. It should be noted that most universities used internal performance based systems for the distribution of research funds, which makes it hard to pinpoint both the time and the strength of the intervention. The CI consisted of external earnings and of output-related indicators: scholarly publications by staff, and the number of higher degrees granted. The weights given to the different elements varied over time, e.g. the

Please cite this article in press as: van den Besselaar, P., et al. Perverse effects of output-based research funding? Butler's Australian case revisited. *Journal of Informetrics* (2017), http://dx.doi.org/10.1016/j.joi.2017.05.016

publication indicator was reduced from 12.5 % to 10 % after an audit that indicated severe problems with data accuracy (Butler, 2003b, p. 40). With hindsight it seems important that the introduced system was not a journal based system only, and definitely not a system solely based on ISI or what is today Web of Science (WoS). Unfortunately this was the image that dominated the subsequent discussion, probably because the Butler papers were based on an analysis of data mainly from the Science Citation Index (SCI).[1]

3. Effects of the new policy – but when?

When exactly the new policy was introduced, and when is it reasonable to expect any type of effect from the "formula-based publication counts"? Butler was not precise about what was exactly the change, and about when the change was implemented (Butler, 2003a; the main article out of a series of papers). Australian research had been losing ground during the 1980s (Bourke & Butler, 1993). Despite the continuing feeling of crisis in the academic community during the 1990s, the publication trends found by Bourke & Butler indicated a positive trend. The Australian share of world publications started to rise and citation levels were regained. In 2003, Butler came back to the question and tried to deconstruct data in order to explain what had happened. The main explanation put forward was pressure on the academic system: declining resources, deterioration of student-staff ratio, and increased performance evaluation.

Rhetorically, Butler asked whether it could be the case that increased 'surveillance' gave researchers an impetus to publish in international journals: As mentioned above, 'publications collections' were introduced in the beginning of 1990s. Since 1992 (or 1993 according to Butler 2002) universities were supposed to supply details of their publication output – collections – to the Ministry. The more specific intervention, funding partly based on numbers of international peer reviewed publications, was not introduced until 1995 (Butler, 2002, 2003b, 2004) and it is unclear whether the actual implementation in the form of governmental funding was set in place before Fiscal Year 1996.

Consequently, the exact moment when changes in the funding regime in Australia would start to affect the publication output is difficult to determine. One can assume that the demand for publications was understood, but because what counts as a publication was not defined, and its effect on funding was unclear, the incentive should be considered slow. Also, there is a time lag between the moment of the decision about an incentive, and the moment when it actually has an effect. To some extent, researchers may change publication channels quickly. But one should expect a minimum time lag of approximately a year (at least), due to the time it takes between submission and publication. Furthermore, when an intervention is directed at universities, it takes time to affect behavior at the individual researcher level. One can also expect effects from the internal university evaluation systems, but it is hard to judge when those were exactly introduced. Overall, one can expect some effect of changes in the funding systems shortly before the mid-nineties, but the *full impact* of changes cannot have occurred before 1998, which is the *last year* in Butler's analysis (see Fig. 1 below). It is evident that Butler's study would have improved greatly if she had used a time series of a couple of years longer – as the findings would have been different. She may not have realized this, as she drew the line between before and after the intervention in 1992 or 1993 (2004,Fig. 17 .1, page 395); but then the universities only had to inform the ministry about publications output, without any funding effect.[2] The effect of output on funding started only after 1995, when the research funding formula was introduced.

4. A discussion about evaluation regimes

Butler's article became important, as it was the first attempt to perform an empirical test of the hypotheses that were formulated by Aldo Geuna and Ben Martin and their colleagues as reactions to the new evaluation regime in the UK. What were these hypotheses? Geuna and Martin (2001, 2003) recognized that they didn't know much about actual consequences: "At present", they wrote, "with the possible exception of the UK, we lack the detailed data on university research inputs and outputs over a sufficiently long period of time needed to identify with confidence the results of different resource allocation systems. Nonetheless, some preliminary thoughts on the advantages and shortcomings of the two approaches (performance based vs. educational size) to university research funding are presented here." Without empirical data, the authors described *potential* disadvantages like the following:

- 'Homogenization' of research and decrease of diversity.
- Discouragement of innovative and risky research.
- Encouragement of publication inflation, e.g. salami publishing and game playing.
- Encouragement of traditional academic research at the expense of societal needs research.
- Separation of research and education, lower priority for teaching.
- Reinforcement of the academic elite, overconcentration of resources.

[1] Misrepresentation of this continues to spread, see for example Hicks (2013) , who argues that the Australian CI "simply counted papers indexed in the WoS" (Cronin & Sugimoto, 2015, p. 671).

[2] Schneider et al. (2016) do the same when discussing the Australian case in their replication.

Please cite this article in press as: van den Besselaar, P., et al. Perverse effects of output-based research funding? Butler's Australian case revisited. *Journal of Informetrics* (2017), http://dx.doi.org/10.1016/j.joi.2017.05.016

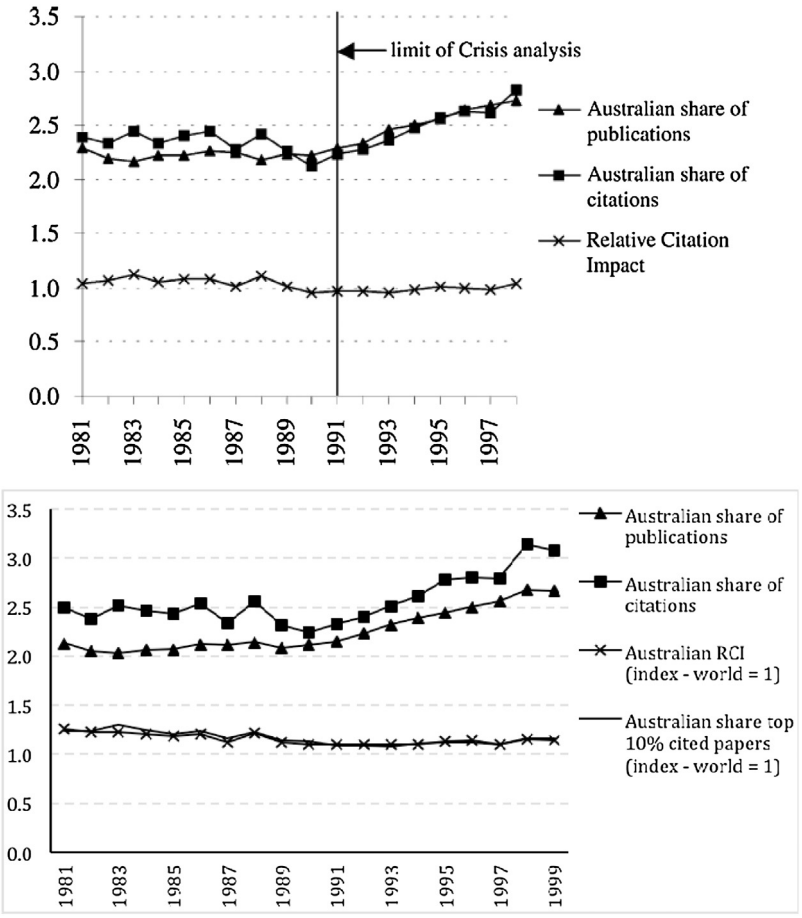

Fig. 1. Australian share in publications, citations and top 10% cited papers (only lower figure) and Australian RCI (upper figure) and NCI (lower figure). Upper figure is copy of Fig. 1 in Butler (2003a). Lower figure is our recalculation of the data using InCites.

Benefits of evaluation systems could not outweigh the costs, which in their mind would be substantial. The authors concluded that all these evaluations were doomed: "In time, the weight of evidence would presumably lead to the assessments being discontinued." (Geuna & Martin, 2001, p. 32). Similar arguments can be found elsewhere (e.g., Gläser, Laudel, Hinze, & Butler, 2002; Weingart, 2005; Whitley, 2007). According to Weingart (2005) "Bibliometric indicators have become such a powerful tool in the context of science policy making and budgetary decisions that *their potentially misleading* and even destructive use must be acknowledged" (italics added). And Whitley (2007) Whitley's (2007) analysis of the consequences of establishing research evaluation systems for knowledge production in different countries and scientific fields is a very interesting hypothesis about possible consequences of research evaluation (and funding) systems, but not an empirical contribution. More recently de Rijcke, Wouters, Rushforth, Franssen, and Hammarfelt (2015) reviewed the literature on the (perverse) effects of evaluation systems on various aspects of the science system. Like in all publications on this topic, the Australian case has a prominent role as probably the only systematic quantitative evidence. And also here, we read for the rest mostly about *possible* effects, and not much about the *prevalence* of (perverse) effects.

As mentioned, Butler was the first to try to test some of the proposed hypotheses with data. In a *Research Policy* article she claimed to be able to explain the effects of formula based funding on publication counts (Butler, 2003a). In a chapter in the *Handbook of Quantitative Science and Technology Research* (Moed et al., 2004) Butler phrased the issue more radically: "What happens when funding is linked to publication counts?" (Butler, 2004), and started taking a position against what others would see as a positive change in the Australian system: the visible hand that directed research towards international journals and away from a local and regional orientation. According to Butler, publishing in ISI-indexed journals would "obviously be the easiest course of action to take" (Butler, 2004, p. 393) as there would in that case be no discussion on whether the selected (WoS indexed) journals were peer refereed or not. With others, she criticized 'quantity oriented policies' as it may result in salami publishing: cutting papers to the smallest publishable entity, which would increase the number of publications but decrease the overall quality and impact (Gläser et al., 2002). We will return to these issues below.

The idea that using productivity in research evaluation – especially in relation to international peer reviewed journal articles – may have as perverse effect the decrease of quality has become a dominant narrative about the science system (e.g.,

Please cite this article in press as: van den Besselaar, P., et al. Perverse effects of output-based research funding? Butler's Australian case revisited. *Journal of Informetrics* (2017), http://dx.doi.org/10.1016/j.joi.2017.05.016

Table 1
Indicators for comparing Australia with the world and with the ten countries.

- Number of publications per country per year (P)
- Number of publications per country per year (P)
- Number of top 10% cited publications per country per year
- Citation impact
  - Relative Citation Impact (RCI) used by Butler: C/P per county per year
  - Normalized Citation Impact (NCI) used by us: It is a subject category and publication type normalized alternative for the RCI – and included in InCites. Details about the calculation are in Appendix A

Hicks et al., 2015; Stephan, 2012; Wilsdon et al., 2015), and the studies of Butler about Australia have become a core element of it. The formula based funding system was conceived bad for Australia, as it would have led to a decreasing relative citation impact (Butler, 2002). This was supported by the observation that the *share* of publications in high impact journals decreased, indicating a change in publication tactics: In Butler's own words: "academics are quick to respond" (Butler, 2010). It does however not seem unlikely that the average JIF of publications in a country would tend to decline, when scientists switch from local mostly not peer reviewed publications to international peer reviewed publications. Such a change of publication tactics would on the short term likely result in publications in less cited journals and therefore decrease the average journal impact of Australia's national output.

Nonetheless such a change would not be detrimental to quality; we actually suggest that such a switch would secure quality as it introduces a quality threshold and enhances quality through the peer review process. Australian researchers became more involved in the international scientific communities and became more focused on communicating results to international audiences. Moreover, publications in journals indexed by the ISI gave higher international visibility to Australian research, something that was explicitly asked for in the discussions during the late 1980s and the beginning of the 1990s. The interesting question, however, is not about the changes in the average journal impact (to which we return later in this paper) but whether the contribution of Australia to the progress of science did increase or not. We suggest measuring this through the number of highly cited papers, and we will investigate whether this indicator decreased or increased after the changes in the funding system.

Several papers have confirmed Butler's findings regarding the increased productivity of the academic sector: Marinova and Newman (2008) concluded that there was ". . .a clear indication that research productivity of the Australian universities has been increasing consistently". Using publications per 100,000 people as indicator, Australia actually did better than UK and New Zealand. While others (Beerkens, 2013; Worthington & Lee, 2008) have restricted the analyses to paper productivity, Marinova and Newman (2008) also found a considerable increase in citations.[3] Recently, Schneider et al. (2016) confirmed part of Butler's findings, notably the lowering of the average impact of journals, using instead of JIF a method for normalization of a journal's impact to the mean of the journal categories where Australian researchers published their papers after the change of funding system.

5. Data and method

To analyze Australia's publication output, we have used data from InCites®, an analysis tool using the data in WoS. The tool provides a number of common bibliometric indicators on subsets from the database, like countries. The definition and calculation of the indicators are straightforward (Table 1) and are explained in Appendix A.

We downloaded from InCites® (Sept 15, 2016) the number of publications, the number of citations, the Normalized Citation Impact (NCI), the number of top 10% cited publications, and the number of non-cited publications. We included only articles, reviews, letters and notes, and not proceeding papers; in that way we do not include the proceedings databases that are not well suited for calculating citation-based indicators. The download was done for the world, for Australia and for ten reference countries mentioned below. We present in this paper the counts for Australia, and most of the analysis consists of comparing Australia with the world, and comparing Australia with the reference countries as selected by Butler.[4]

The main drawback in using InCites® is that it uses full counts only: any article with at least one Australian address is counted as one full Australian article. Since co-authorship between institutions and countries increases over time, full counting induces trends in publication numbers that can be attributed to differences in the rate of co- authoring. Small countries cannot be compared to large, and a country's subset can be expected to increase relative to the world output. As we selected comparable reference countries, these effects do not seriously influence the analysis. As far as it does, our study

[3] That part was based on two case studies, one on their own institute and one using Google Scholar.
[4] As the data are proprietary, we cannot publish those. The data can easily be retrieved if one has access to an InCites account.

Please cite this article in press as: van den Besselaar, P., et al. Perverse effects of output-based research funding? Butler's Australian case revisited. *Journal of Informetrics* (2017), http://dx.doi.org/10.1016/j.joi.2017.05.016

will in this respect suffer from the same problems as Butler's study, as Butler (2003a) used National Science Indicators (NSI) to compare Australia to other countries. That tool can be regarded as the precursor of InCites® and also uses full counts.

InCites® deploys an open citation window[5] and this will result in differences compared to Butler's data, especially towards the end of the nineties. As our citation data are based on WoS, September 2016, and the analysis by Butler was based on data close to the end of the millennium, our citation counts are higher. However, our more recent data are more precise and more valid because of the longer citation windows. Furthermore, whether the relative citation impact increased or decreased is difficult to observe, as the scale used by Butler is simply unsuitable for observing changes in the variable. Our data also differ because of changes in WoS, which has increased the number of indexed journals. As a consequence of that, Australia's share of world publications becomes significantly lower in our data (Fig. 1). The variation in publications also affects the relative citation impact, which in our data is slightly higher.

Regardless of the differences, the two figures convey approximately the same message as to how Australian publication output varied during the 80s and 90s. Up to the early 1990s, Australia's share in publications and citations was stable, and went up during the rest of the 90s. The recovery may however be the result of increased international co-authoring, which, since co-authoring will not have any effect on world counts, will inflate the Australian share of world publications, citations and impact (the problem is described well in Butler, 2003a). To avoid this problem we have chosen to compare Australia with a set of similarly developed countries. We have used the same countries as Butler selected for comparison: Belgium, Canada, UK, France, Germany, Italy, Japan, Netherlands, Sweden and Switzerland (plus Australia).

As the main issue in this paper is the development of impact as a result of the changed research policy, we use the number and share of *top 10% cited papers* within each WoS subject category as indicators. Using the top 10% cited papers has two advantages over using the *Relative Citation Index* (RCI), as used by Butler. The RCI is calculated as "dividing Australia's share of world citations by its share of world publications" (Butler, 2003a). Why do we prefer the top 10% cited papers? Firstly, this indicator yields a direct measure of impact, while the RCI (and the Incites® variant of it, the NCI) is dependent on variations in low or non-cited papers, papers that do not contribute to impact. Another advantage of using the top 10% cited papers is that this indicator is less sensitive for outliers: One or a few extremely highly cited papers can strongly increase the RCI.[6]

At the behavioral level, we do three analyses. We firstly use the number of non-cited papers to investigate possible changes in publication tactics. This also sheds light on the changes in the average JIF of Australian publications as reported by Butler (2004). Secondly, in order to follow in which types of journals (regional or international) Australian researchers published, we draw a subsample of the 170 journals with most Australian addresses. From this subsample we classified journals that have 'Australia' in their name as 'regional', and we classified the rest as 'international'. Thirdly, we briefly test whether the salami strategy was deployed. If Australian researchers did use this strategy, we would expect them having relatively short articles compared to other countries, and we would expect that some years after the implementation of the new 'quantity policy' a downward trend can be observed in the Australian paper length.[7] At least, we do not expect that reducing the scientific content would leave the length of papers untouched.

6. Policy changes and Australian publications 1981–2007

As in most developed countries, Australian publication output has been increasing every year with the exception of 1989, as did the number of top 10% cited papers – and from the mid-1990s onwards faster than total publication output (Fig. 2).

How did Australia do compared with the world, and especially compared with the reference set? For the comparison with the world, we divide for each year the number of (citable) publications with an Australian address by the total number of (citable) papers (as was done for Fig. 1), and this results in the dotted grey line in Fig. 3. We do the same, for the top 10% cited papers. Compared to the world, Australian share in world production shows a very small decline from 1982 to 1991 (Fig. 3), followed by an immediate increase that continues up to 2011. This suggests, as Butler pointed out, that there was a very quick response to the discussion of crisis and the policy changes that occurred immediately thereafter. Also the impact – measured as the *share* of world top 10% cited papers – started to increase, after a decade of a small decline between 1982 and 1992 (Fig. 3, black dotted line).

As it is more meaningful, we also compare the Australian share in output and in top 10% cited papers with the reference countries (Fig. 3, straight lines). This was done in the following way: We calculate for each year the total number of publications N of the reference countries plus Australia.[8] N(Aus)/N gives the share of Australian papers. We do the same for the number of top 10% cited papers. As Fig. 3 shows, compared with the reference countries, the pattern is more pronounced. Relative productivity (the black dotted line) and impact (black straight line) decline steeply until the mid-1990s, but recover strongly from 1995 onwards. This clearly contradicts Butler's claim of a decreasing impact compared to the reference countries. As we have argued above, the full impact of policy changes cannot be expected before around 1998 and indeed impact

[5] Open citation window: all citations are included until the moment of the download of the data. This means that older papers have had more time to be cited than newer publications.

[6] For a further discussion on the choice of indicators, see e.g. Tijssen et al. (2003), and Waltman and Van Eck (2015).

[7] A few years after, as it takes a while before this becomes visible: papers need firstly to be written, and then it takes time for reviewing and publishing.

[8] As the database does not fractionalize, papers with authors in several reference countries are counted multiple times. This is similar to Butler's method, using the ISI National Science indicators (NSI).

Please cite this article in press as: van den Besselaar, P., et al. Perverse effects of output-based research funding? Butler's Australian case revisited. *Journal of Informetrics* (2017), http://dx.doi.org/10.1016/j.joi.2017.05.016

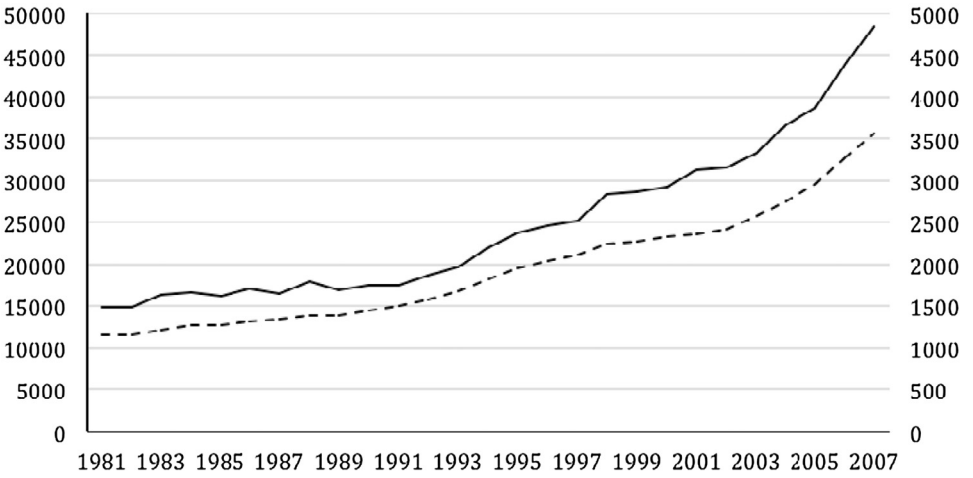

Fig. 2. Australian output (articles, reviews, letters and notes only; dotted line; left axis) and number of top 10% cited papers (straight line; right axis);.
Source: InCites

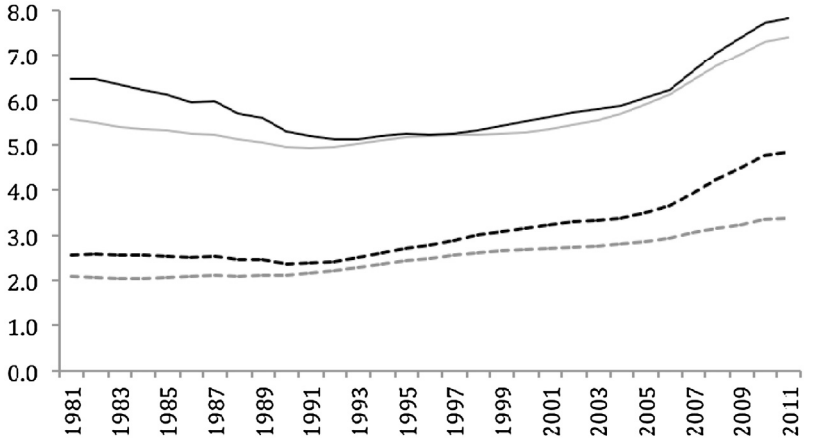

Fig. 3. Australian output and impact.
Australian output as percentage of world output (dotted grey line). Australian top 10% cited papers as percentage of world top 10% cited paper (dotted black line). Australian output as percentage of the reference countries (straight grey line). Australian top 10% cited papers as percentage of the reference countries (straight black line)
Source: Calculated using data from InCites

starts to increase rapidly from that year onwards. The increased evaluation practices and the introduction of funding based on publication counts evidently coincide with markedly increased impact – also when compared to the reference set of countries selected by Butler.

In Fig. 4, we plot next to the top 10% cited papers also the NCI, similar as in Fig. 1, but now using an appropriate scale (dotted grey line). Obviously, that indicator also goes up – actually already since 1992–1993. It is visible in this graph as we use – in contrast to Butler (see Fig. 1) – an appropriate scale. Comparing Australia with the selected countries, we see a modest increase of the NCI, after a decade decline until 1995. In all the three time series, there is a break of the negative trend.

The share of top 10% cited papers is also useful for comparing the research impact (as proxy for quality) between countries. In Fig. 5 we show the relative score for the eleven countries: If a country has a score of 1.30, it means that in that year the country has a 30% higher share of top 10% cited papers than the average of the set of eleven countries. The emerging picture is rather clear. In the beginning of the 1980s, the countries were divided in one highly cited group (Sweden, Netherlands, UK, Canada, Australia and Switzerland) and one less cited group (Belgium, France, Germany Italy and Japan). Since then Netherlands and Switzerland have kept their high share of top 10 % cited publications, whereas the four other countries show a decline. Of those four, only Australia and possibly Sweden have reversed the trend.

In the low cited group all countries except Japan have increased their share of top 10% cited publications and thereby of course contributed to Australia's relative decline. The picture suggests a substantial convergence in the world science system, with some exceptions at the higher end (Netherlands, Switzerland) and at the lower end (Japan). Australia, after the 1995- change in the funding system, is doing relatively well from 1997 onwards. It should be noted that the curve does not

Please cite this article in press as: van den Besselaar, P., et al. Perverse effects of output-based research funding? Butler's Australian case revisited. *Journal of Informetrics* (2017), http://dx.doi.org/10.1016/j.joi.2017.05.016

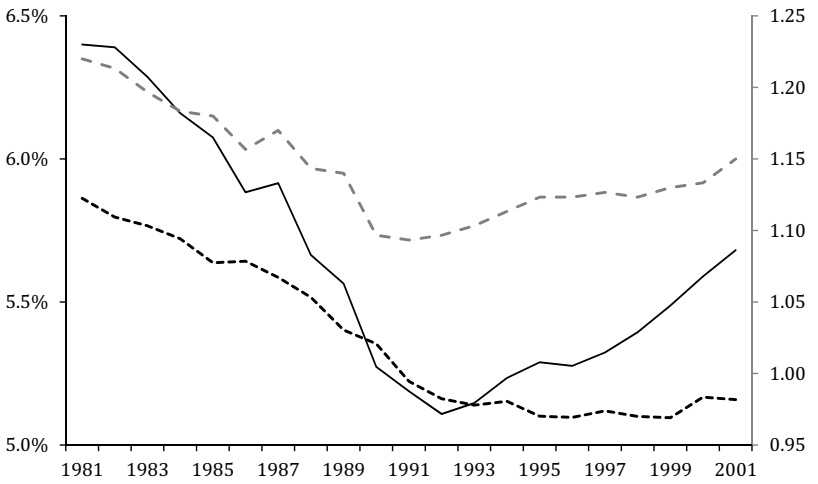

Fig. 4. Change in the Australian impact.
Straight black line: Australian top 10% cited papers, share of the eleven countries (left axis). Dotted black line: Australian NCI, relative to the average NCI of the eleven countries (right axis). Dotted grey line: Australian NCI, relative to world average (right axis)
Source: Calculated using data from InCites

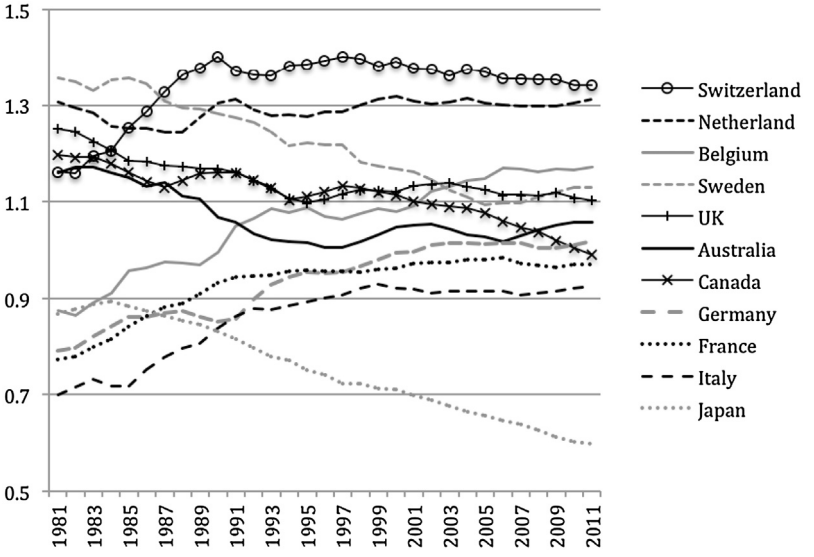

Fig. 5. Australia's share of top 10% cited papers compared to the share of the reference countries.
Share for reference countries = 1; 1.3 means that share is 30% above average; 3 years moving average.
Source: Calculated using data from InCites

imply a slow development of impact: it shows that Australia's *share* of top 10% cited papers goes up slowly, but as Australia's total number of papers increases rapidly (Fig. 2), this also holds for Australia's *absolute number* of top 10% cited papers.

7. Publication strategies

To explain her observations, Butler also looked at the distribution of Australian publications over JIF percentiles. She argued that Australian researchers adapted their publication strategy in response to the new output based funding policy by producing more articles and by publishing these articles increasingly in the lower JIF quartiles – suggesting that the larger production was of lower quality and not suited for publishing in top journals. This is a problematic argument, as citation impact and output quality is a characteristic of a paper, but the JIF is an attribute of journals. There is a correlation of course, but not very strong (Seglen, 1997). Recently a study showed that the most cited articles are published less exclusively in high JIF journals (Lozano et al., 2012).

Please cite this article in press as: van den Besselaar, P., et al. Perverse effects of output-based research funding? Butler's Australian case revisited. *Journal of Informetrics* (2017), http://dx.doi.org/10.1016/j.joi.2017.05.016

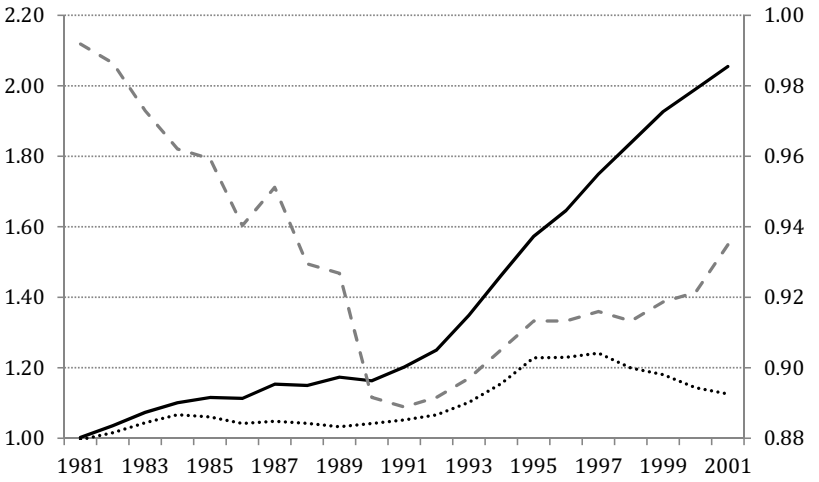

Fig. 6. Australian top cited publications, non-cited publications, and NCI.
Straight black line: Australian top 10% cited papers; index 1981 = 100 (left axis). Dotted black line: Australian non-cited papers; index 1981 = 100 (left axis). Dotted grey line: Australian NCI, relative to world average (right axis)
Source: Calculated using data from InCites

Nevertheless, an increased number of high impact papers could be accompanied with an even larger increase of lowly cited papers (in low impact journals), and together result in a decreasing average impact.[9] We therefore plot (Fig. 6) the development of the non-cited papers, the highly cited papers, and the NCI, covering the period up to 2010.[10]

The number of *top 10% cited papers* did not decrease during the period of 'crisis' but instead develops at a slow and steady pace. After 1996 the increase of Australian share of highly cited papers accelerates, (and as shown in Fig. 3, faster than in the reference counties). On the other hand, the number of *non-cited papers* remains stable until 1991, and then increases slowly until 1996. Immediately after the introduction of output based funding, the production of highly cited papers accelerates, whereas the number of non-cited papers decreases (Fig. 6). Without suggesting that there is a direct relationship between the three indicators, we see that the growing number of top cited papers, and a decline of non-cited papers, coincides with an increasing NCI. The decline of non-cited papers does *not* suggest that Australian researchers adapted a high volume – low quality strategy.

How does this relate to Butler's observation about the JIF quartiles (Butler, 2002)? The downward trend in the non-cited papers suggests that the papers of Australian authors in low impact journals were cited, and therefore despite the strongly increasing share of low impact journals in the total Australian output, the average impact still increased, as did the number of highly cited papers. In fact, based on the top 10% cited papers, impact increased by about 25% between 1991 and 2001–an increase that indicates the strengthening of Australian research in those days.[11]

So the findings of Butler about the JIF percentiles do not conflict with our findings about the increasing impact of Australian science. Apart from the too short time frame, she interpreted the journal choice incorrectly. It seems as if those researchers who, because of the new regime, started to publish in international (WoS indexed) journals may have submitted to lower impact journals first – although the average impact of those papers proved reasonable. Probably the obsession with the JIF was in those days less dominant compared to the situation nowadays.

We also investigate whether Australian researchers moved to local (often lower impact) journals. Fig. 7 shows the growth of Australian output over the period 1990 until 2005 split into international and local journals (here defined as journals with "Australia*" in the title[12]) – all included in the WoS. Obviously, the growth is mainly in international journals, indicating a reorientation of Australian science towards internationalization.

Finally, we investigate whether Australian researchers started to 'salami slice' their papers, in order to increase output levels. This strategy refers of course primarily to slicing the cognitive content in small parts. However, we hypothesize that if the cognitive content is reduced, one will also see a relative reduction of the length of papers, in order to avoid that reviewers evaluate papers negatively ("many pages, little content"). If one may use reduction of the average number of pages

[9] We already showed that this decrease of average citation impact did not take place. But it is still interesting to go into this issue, in order to cover the various arguments by Butler.
[10] Using non-cited papers with an open citation window makes the probability of being non-cited time dependent: older publications have more time to be cited than more recent papers. As we only take into account 1981-2001, we assume that this effect is taken away.
[11] The number of top 10% cited papers increased between 1991 and 2001 by 41%. Correcting for an increase in international co-authorship of about 15% in the same period, leads this to about 25% 'real' increase.
[12] We tested also to use the address of the publishing house (PA tab in WoS) and obtained a similar picture.

Please cite this article in press as: van den Besselaar, P., et al. Perverse effects of output-based research funding? Butler's Australian case revisited. *Journal of Informetrics* (2017), http://dx.doi.org/10.1016/j.joi.2017.05.016

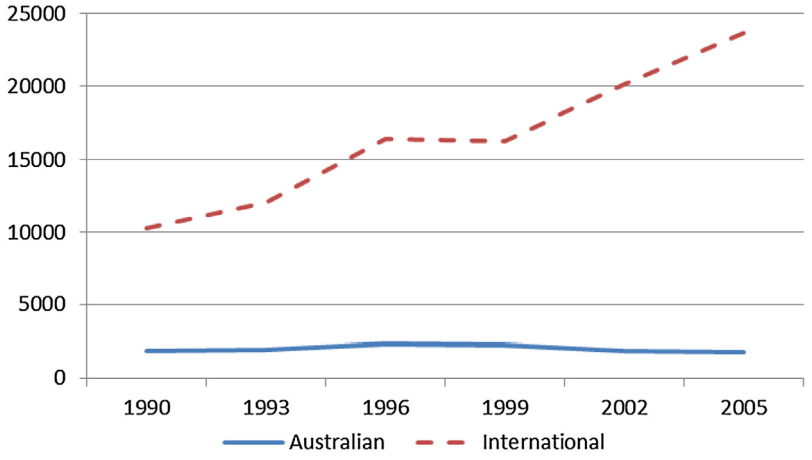

Fig. 7. Australian output (articles only) in international and local journals 1990-2005.
The strong growth in international output in 1996 is due to an extension of the database (12% increase), indicating that a higher number of journals were covered.
Source: Calculated using WoS data

as indicator for salami slicing (c.f. Broad, 1981), one can test whether that occurred in the period under study. We therefore calculated the (weighted[13]) average page length in twelve journals[14]:

- Inorganic Chemistry
- Journal of Applied Physics
- Monthly Notices of the Royal Astronomical Society
- Tetrahedron Letters
- Blood
- Endocrinology
- Journal of Immunology
- Lancet
- Pathology
- Marine Biology
- Wildlife Research
- Lecture Notes in Artificial Intelligence

The Australian papers do not show a decreasing length after the new policy was implemented, and they are relatively long compared to the selected set of other countries (Fig. 8). In fact, the trend is slightly positive for Australia, as it is for all countries together. This analysis cannot show whether salami slicing took place in Australia in the 1990s. However, as the Australian pattern does not deviate from the overall pattern, we can conclude that salami slicing – if it took place – was similar to the rest of the world, and therefore cannot explain the increasing share of Australia in the world research output.

8. Conclusions and discussion

Australia showed a drop in performance in the beginning of the nineties, in terms of publication share and citation-based quality indicators: the Relative Citation Impact and the share top 10% cited papers. This had been a continuous trend since the beginning of the 1980s, but all parameters show a recovery from the mid-1990s. The recovery is especially strong compared with the whole world, and slightly less strong but positive compared to the selected set of scientifically strong countries. However, in both comparisons there is a break with the downward trend characterizing the 1980s and the early 1990s. Therefore, Butler's conclusions about the effects of the changes in the funding system cannot be maintained.

When comparing individual countries, the decline since the beginning of the 1980s does not seem to be a specific problem for Australia, as we also observe this for Sweden, UK and Canada. It seems an effect of the convergence of research performance of developed countries due to increased international collaboration: through increased international co-authoring, high impact publications are shared between countries. Only a few countries show a different pattern, in positive or negative direction.

[13] This was done to control for the differences in page length between fields.
[14] The selection criterion was that Australia needed to have a fair amount of papers in the journal. Furthermore, we selected journals from different fields.

Please cite this article in press as: van den Besselaar, P., et al. Perverse effects of output-based research funding? Butler's Australian case revisited. *Journal of Informetrics* (2017), http://dx.doi.org/10.1016/j.joi.2017.05.016

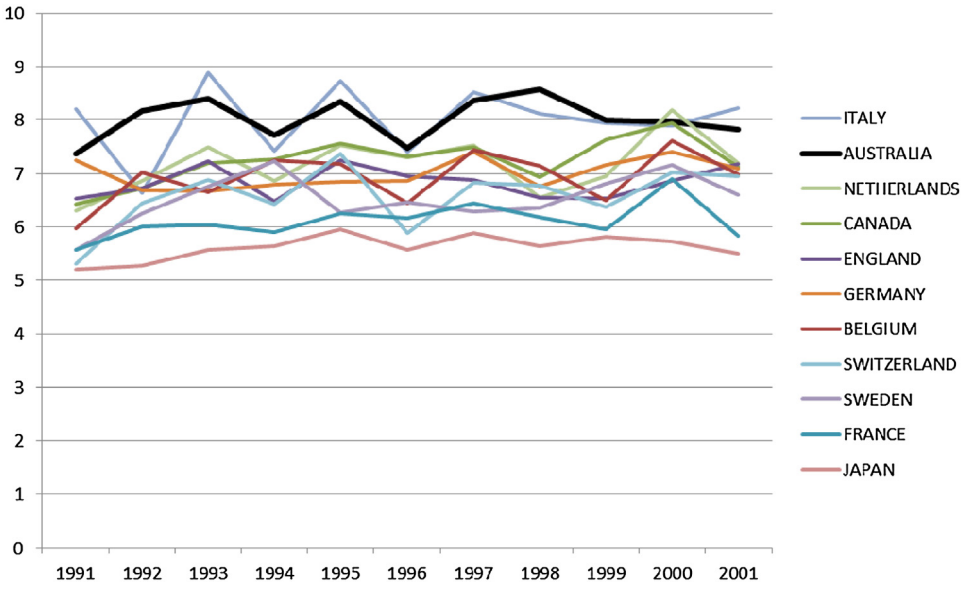

Fig. 8. Average page length (number of pages) per year by country.
Source: WoS (2016-04-06).

At the behavioral level, the changes in publishing strategies of Australian researchers did not result in publishing more in local and regional peer reviewed journals, but in international ones. And we also do not find evidence that Australian academics started to salami slice papers, in order to increase output.

We should refrain from drawing *too stark* conclusions regarding the causality of the process described (Osuna et al., 2011). In the case of national research systems (or national university systems) there are many other factors, some of which Butler included in her discussion: e.g. sector differences, funding changes, personnel changes. Other relevant factors were not taken into account, such as the main changes taking place in the university sector in the Western world at the end the 1980s and the beginning of the 1990s. Competition for grants from research councils became tighter, which in itself made researchers underscore the value of international publications, i.e. the core of the academic self-understanding. And, evaluation based on journal publications was increasingly a common phenomenon at that time (Evered & Harnett, 1988). Most probably, improving productivity does not only depend on introducing output based funding systems, but also on other factors as we suggested elsewhere (Sandström, Heyman, & van den Besselaar, 2014).

Nevertheless, the data do suggest that the new policy during the 1990s gave the Australian science system a new impulse – as funding partly became output-dependent after 1995. This initiative – likely together with other changes in the system mentioned above – did not only contribute to higher productivity, but – as can be expected from creativity theory (Simonton, 2004) – also to higher quality. Using a longer time series and using scales that match the variation of the variables, we show a positive and not a negative relationship between the size of the output and its quality. Quantity matters, not only at the level of individual researchers (Larivière & Costas, 2015; Sandström & van den Besselaar, 2016; Van den Besselaar & Sandström, 2015) but also at the level of the science system.[15] If researchers are stimulated to become more productive and to produce more ideas (and papers), they on average produce a higher proportion and a higher absolute number of good ideas (and of good papers). This is an important finding for science policy and for research management.

What do we learn from this for recent discussions about the use of indicators for research assessment (Hicks et al., 2015; Wilsdon et al., 2015; de Rijcke et al., 2015)? That debate focuses on four issues: (i) quality of the indicators; (ii) performance dimensions covered; (iii) the relation between indicators and peer review; (iv) perverse effects of indicator use. Our findings stand in strong contrasts to many arguments that there is a too strong emphasis on measuring productivity (first issue). It also contradicts popular arguments that emphasis on productivity has perverse and negative effects on the quality of science (last issue). Both arguments refer to a single empirical source: the Australian case. As these arguments are increasingly taken up by evaluation systems, and e.g., allow researchers to submit only a few papers, the perverse effect may be that the better – and often highly productive – researchers are disadvantaged: they are not allowed to show their full work to the evaluators.

A second lesson is about replication studies. Butler's negative evaluation of the Australian funding system has been uncritically used in many other studies (Hicks, 2009, 2012; OECD, 2010; Schneider et al., 2016; de Rijcke et al., 2015). This

[15] And at the level of universities: within the Leiden Ranking, there is also a positive relation between total output at the university level and total and share of top 10% cited papers.

Please cite this article in press as: van den Besselaar, P., et al. Perverse effects of output-based research funding? Butler's Australian case revisited. *Journal of Informetrics* (2017), http://dx.doi.org/10.1016/j.joi.2017.05.016

has had consequences for scholarly work[16] and for policies. But science policy researchers have refrained from checking the data and from replication. The same holds for the discussion on 'perverse effects' of quantitative performance indicators, such as on slicing of papers. These phenomena may occur, but convincing empirical support is lacking. We showed some results suggesting the opposite, but of course more studies are needed. For the moment, we conclude that much of the criticism on performance based funding and its effects seem strongly overstated and even simply wrong.

Acknowledgements

The work underlying this paper was funded by the Riksbankens Jubileumsfond, grant P12-1302:1. We thank three reviewers for criticizing an earlier version of this paper, and the editor of this journal for detailed comments.

Appendix A.

The InCites indicators

We here describe how the deployed indicators are calculated by InCites, summarizing the information provided by the InCites indicators handbook. It was downloaded from http://researchanalytics.thomsonreuters.com/m/pdfs /indicators-handbook.pdf as of December 28, 2016.

InCites is based on the seven components of the WoS Core Collection:

- Science Citation Index Expanded,
- Social Sciences Citation Index,
- Arts & Humanities Citation Index,
- Conference Proceedings Citation Index – Science,
- Conference Proceedings Citation Index – Social Sciences & Humanities,
- Book Citation Index – Science
- Book Citation Index – Social Sciences & Humanities.

As we only included articles, review, letters, and notes, we cover only the first three databases. This is necessary if one wants to use citation-based indicators.

Subject areas of InCites follow the WoS categorization into 252 subjects areas. Also, papers in multidisciplinary journals (e.g. Nature, Science) are reclassified to their most relevant subject areas.

Citation indicators are calculated using a baseline based on the performance of a global set of publications with the same subject area, document type, and publication year. We use two citation-based indicators in this paper.

- The top 10% cited papers. This is calculated as the top 10% most cited documents in a given subject category, year and publication type (document type) divided by the total number of the documents in a given set of documents, displayed as a percentage. A higher value is considered to be higher performance.
- Category Normalized Citation Impact (NCI) of a document is calculated by dividing the actual count of the citing items by the expected citation rate for documents with the same document type, year of publication and subject area. When a document is assigned to more than one subject area, an average of the ratios of the actual to the expected is used. The NCI of a set of documents, e.g. the collected works of an individual, institution or country, is the average of the NCI values for all the documents in the set.

For a single paper that is assigned to only one subject area, this can be represented as:

$$NCI = c/e_{ftd}$$

For a single paper that is assigned to multiple subject areas, the NCI can be represented as the average of the ratios of actual to expected citations for each subject area:

$$NCI = (\Sigma_j c/e_{f(j)td})/n$$

For a group of papers, the NCI value is the average of the values for each of the papers, represented as:

$$NCI = (\Sigma_j \mathrm{NCI_j})/p$$

[16] An example is the comparison of the 'Norwegian model' with the Australian case by Schneider et al. (2016). The Norwegian model and the Australian model had similar positive effects on productivity. The authors claim that the Norwegian model goes together with better impact development (no change) than the Australian (decline). This conclusion is obviously based on Butler's account of the Australian case. In fact, the Norwegian model taking into account also the *quality* of output performed worse in terms of impact than the Australian system, which did not take quality into account (Van den Besselaar & Sandstrom, forthcoming).

Please cite this article in press as: van den Besselaar, P., et al. Perverse effects of output-based research funding? Butler's Australian case revisited. *Journal of Informetrics* (2017), http://dx.doi.org/10.1016/j.joi.2017.05.016

Where:

- e = expected citation rate or baseline
- c = times cited
- p = number of papers
- f = field or subject area
- t = year
- d = document type
- n = number of subject areas a paper is assigned to

References

Bourke, P., & Butler, L. (1993). *A crisis for Australian science? Canberra: Research school of social sciences, ANU – monograph series*. Australian National University [Performance Indicators Project; no.1]
Broad, W. J. (1981). The publishing game: Getting more for less. *Science*, *211*, 1137–1139. http://dx.doi.org/10.1126/science.7008199
Butler, L. (2002). A list of published papers is no measure of value: The present system rewards quantity, not quality—But hasty changes could be as bad. *Nature*, *419*, 877.
Butler, L. (2003a). 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. (2003b). Modifying publication practices in response to funding formulas. *Research Evaluation*, *12*(1), 39–46.
Butler, L. (2004). What happens when funding is linked to publication counts? In H. F. Moed, W. Glanzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology* (pp. 340–389). London: Kluwer Academic Publishers.
Butler L. (2010) Impacts of performance-based research funding systems: A review of the concerns and the evidence. In OECD (2010), *Performance-based Funding for Public Research in Tertiary Education Institutions:* Workshop Proceedings, pp. 127–165.
Cronin, B., & Sugimoto, C. R. (2015). *Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact*. Cambridge, MA: MIT Press.
de Jong, S., Barker, K., Cox, D., Sveinsdottir, T., & Van den Besselaar, P. (2014). Understanding societal impact through studying productive interactions. *Research Evaluation*, *23*(2), 89–102.
de Rijcke, S., Wouters, P. F., Rushforth, A. D., Franssen, T. P., & Hammarfelt, B. (2015). Evaluation practices and effects of indicator use—A literature review. *Research Evaluation*, *25*(2), 161–169.
Evered, D., & Harnett, S. (Eds.). (1988). The evaluation of scientific research. In *Ciba foundation conference*. Chichester: John Wiley.
Geuna, A., Martin, B. (2001). *University Research Evaluation and Funding: An International Comparison*. SPRU Electronic Working Paper Series No. 71.
Geuna, A., & Martin, B. (2003). University research evaluation and funding: an international comparison. *Minerva*, *41*, 277–304.
Gläser, J., Laudel, G., Hinze, S. & Butler, L. (2002). *Impact of evaluation-based funding on the production of scientific knowledge: What to worry about, and how to find out*. (Expertise for the German Ministry of Education and Research). [Available from Research Gate accessed 2016-04-06].
Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: the leiden manifesto for research metrics. *Nature*, *22*, 2015.
Hicks, D. (2009). Evolving regimes of multi-university research evaluation. *Higher Education*, *57*, 393–404.
Hicks, D. (2012). Performance-based university research funding systems. *Research Policy*, *41*, 251–261.
Hicks, D. (2013). One size doesn't fit all: on the co-evolution of national evaluation systems and social science publishing. Confero: essays on Education. *Philosophy and Politics*, *1*(1), 67–90 [(reprinted in (Eds.) Cronin & Sugimoto, *Scholarly Metrics under the Microscope*. New Yersey: Information Today Inc. 2015, pp. 661–677).
Larivière, V., & Costas, R. (2015). How many is too many? On the relationship between output and impact in research. *PLoS One*, *11*(9), e0162709. http://dx.doi.org/10.1371/journal.pone.0162709
Lozano, G. A., Larivière, V., & Gingras, Y. (2012). The weakening relationship between the impact factor and papersè citations in the digital age. *Journal of the American Society for Information Science and Technology*, *63*(11), 2140–2145. http://dx.doi.org/10.1002/asi.22731
MacRoberts, M. H., & MacRoberts, B. R. (1989). Problems of citation analysis—A critical review. *JASIS*, *40*(5), 342–349.
Marinova, D., & Newman, P. (2008). The changing research funding regime in Australia and academic productivity. *Mathematics and Computers in Simulation*, *78*(2–3), 283–291.
Martin, B. R., & Irvine, J. (1983). Assessing basic research—Some partial indicators of scientific progress in radio astronomy. *Research Policy*, *12*(2), 61–90.
Narin, F. (1976). *Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity*. Computer Horizons Inc [456 pp].
Nederhof, A. J., & van Raan, A. F. J. (1993). A bibliometric analysis of 6 economic research groups—A comparison with peer review. *Research Policy*, *22*(4), 353–368.
OECD. (2010). *Performance-based funding for public research in tertiary education institutions: Workshop proceedings*. OECD.
Osuna, C., Cruz-Castro, L., & Sanz-Menendez, L. (2011). Overturning some assumptions about the effects of evaluation systems on publication performance. *Scientometrics*, *86*, 575–592.
Pelz, D. C., & Andrews, F. M. (1966). *Scientists in organizations: Productive climates for research and development*. New York: Wiley.
Phelan, T. J. (1999). A compendium of issues for citation analysis. *Scientometrics*, *45*(1), 117–136.
Roche, T., & Smith, D. L. (1978). Frequency of citations as criterion for the ranking of departments, journals and individuals. *Sociological Inquiry*, *48*, 49–57.
Sandström, U., & van den Besselaar, P. (2016). Quantity and/or quality? The importance of publishing many papers. *PLoS One*, *11*(11), e0166149. http://dx.doi.org/10.1371/journal.pone.0166149
Sandström, U., Heyman, U., & van den Besselaar, P. (2014). The complex relationship between competitive funding and performance. In *Conference: Science technology indicators*.
Schneider, J. W., Aagaard, K., & Bloch, C. W. (2016). What happens when national research funding is linked to differentiated publication counts? A comparison of the Australian and Norwegian publication-based funding models. *Research Evaluation*, *25*(3), 244–256.
Seglen, P. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ*, *314*, 497.
Simonton, D. K. (2004). *Creativity in science: Chance, logic genius, and zeitgeist*. N.Y. Cambridge Univ Press [Reprinted 2008].
Stephan, P. (2012). Perverse incentives. *Nature*, *484*(212), 29–231.
Van Arensbergen, P., Van der Weijden, I., & van den Besselaar, P. (2014). Different views on scholarly talent—What are the talents we are looking for in science? *Research Evaluation*, *23*(4), 273–284.
Van Raan, A. F. J. (1996). Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. *Scientometrics*, *36*(3), 397–420.
Van den Besselaar, P., & Sandström, U. (2015). Does quantity make a difference? In A. Salah, S. Sugimoto, & U. Al (Eds.), *Proc. ISSI 2015 conference* (pp. 577–583).
Van den Besselaar P., SandstroÅNm U., Counterintuitive effects of incentives? Research Evaluation (Forthcoming).
Van den Besselaar, P., Sandström, U., & van der Weijden, I. (2012). The independence indicator. In E. Archambault, Y. Gingras, & V. Lariviere (Eds.), *Science & Technology Indicators* (pp. 131–141). Montreal: OST & Science Metrix.

Please cite this article in press as: van den Besselaar, P., et al. Perverse effects of output-based research funding? Butler's Australian case revisited. *Journal of Informetrics* (2017), http://dx.doi.org/10.1016/j.joi.2017.05.016

Van der Weijden, I., Verbree, M., & van den Besselaar, P. (2012). From bench to bedside: The societal orientation of research leaders. The case of biomedical and health research in the Netherlands. *Science & Public Policy*, *39*, 285–303.
Van der Weijden, I., Belder, R., van Arensbergen, P., & van den Besselaar, P. (2015). How do young tenured professors benefit from a mentor? Effects on management, motivation and performance. *Higher Education*, *69*, 275–287.
Waltman, L., & Van Eck, N. J. (2015). Field-normalized citation impact indicators and the choice of an appropriate counting method. *Journal of Informetrics*, *9*(4), 872–894.
Wilsdon, J. et al. (2015). The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management. 10.13140/RG.2.1.4929.1363.

Please cite this article in press as: van den Besselaar, P., et al. Perverse effects of output-based research funding? Butler's Australian case revisited. *Journal of Informetrics* (2017), http://dx.doi.org/10.1016/j.joi.2017.05.016