

Counterintuitive effects of incentives?

Peter van den Besselaar,^{1,*} and Ulf Sandström²

¹Department of Organization Sciences & Network Institute, Vrije Universiteit Amsterdam, Amsterdam, Netherlands and ²Department of INDEK, KTH Royal Institute of Technology, Stockholm SE-100 44, Sweden

*Corresponding author. Email: p.a.a.vanden.besselaar@vu.nl.

Abstract

A recent paper in this journal compares the Norwegian model of using publications counts for university funding with a similar intervention in Australia in the mid-1990s. The authors argue that the Norwegian model (taking into account the quality of publications) performs better than the Australian (which did neglect paper quality other than being peer reviewed). We argue that these conclusions are in contrast to the evidence provided in the article, and therefore should be considered incorrect.

A recent paper in this journal argues:

The experiences from Australia where undifferentiated publication counts were linked to funding of universities in 1993 is well known. Publication activity increased, but the largest increase was in lower-impact journals, leading to a general drop in overall citation impact for Australia. The experience from Australia has been a warning for what would most likely happen if funding were linked to publication activity. Nevertheless, in 2005, a performance-based model based on differentiated publication counts was implemented in Norway. The model was specifically developed to counter adverse effects like those identified in the Australian case. (Schneider, Aagaard and Bloch 2016)

Basically, the Norwegian system is counting research output in ‘quality classes’ for two quality levels (Table 1) and four publication types (Table 2). The higher Level 2 includes approximately 8% of the journals and about 5.5% of the publishers of books and proceedings. The criteria for the levels and the weights of the publication types seem arbitrary. The Norwegian model does take into account the number of authors per publication. If three authors publish an article in a Level 2 journal, each author receives 1 point. If the same authors publish in a Level 1 journal, they receive each 0.33 points.

Schneider and his colleagues claim that the system indeed resulted into a growing productivity, without negative effects on

quality (here defined as impact): both the Norwegian system and the Australian system resulted in higher output, but in Australia the citation impact went down, whereas in Norway it remained stable—so the Norwegian system does function better. As the ‘Norwegian model’ is currently under consideration or being implemented in several countries (such as Denmark; Finland; Sweden; Belgium), it is—according to the authors—an important finding as it would show that the Norwegian model is an effective science policy instrument, as it realizes the goals without negative side effects.

However, the evidence presented in the article does not support the authors’ conclusions. The authors have several graphs about Australia in their article, and the core of the evidence is in Fig. 7 (reproduced here as Fig. 1). This figure reveals no declining impact in the Australian case, as immediately after the output-based funding was implemented in 1993,¹ the decline of Australian performance stopped and the impact of the Australian system started to increase again.² This holds especially for the top 10% highly cited papers (Squares), and the less meaningful average impact indicator starts to increase shortly later (Triangles). Following these findings of Schneider et al., we would argue that the article shows that the Norwegian model did not reach its goal (‘specifically developed to counter adverse effects like those identified in the Australian case’): the Australian system not only had a positive effect on output level,

Table 1. Classification of media in the *Norwegian Model in 2015*

	Level 1	Level 2
Journals	22,543	2,042
Publishers	1,484	86

Table 2. The value of publications in the *Norwegian Model in 2015*

	Level 1	Level 2
Journal article	1	3
Chapter in book	0.7	1
Proceedings	0.7	1
Book	5	8

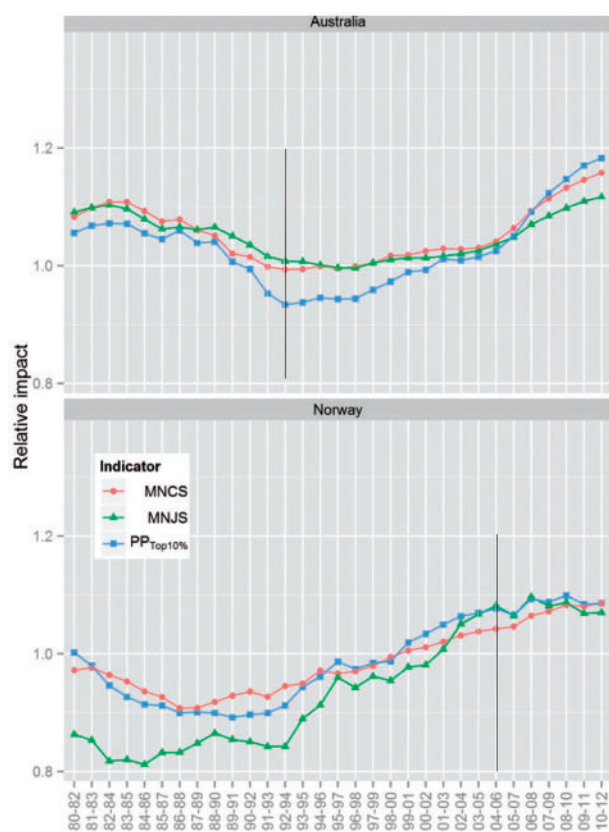


Figure 1. Performance of Australia and of Norway.

Note: Squares: Share top 10% highly cited papers. Circles Mean normalized citation score. Triangles: Mean Normalized Journal Score Figure is adapted from Schneider, Aagaard and Bloch 2016, p 252.

but in contrast to the Norwegian model, it also had a positive effect on impact. Even stronger, both the Australian and the Norwegian policy interventions were followed by a trend reversal—but in different directions. In Australia, the downward impact trend was replaced by an upward trend; whereas after the Norwegian policy intervention, the upward development of impact seems to end.

Why did the authors observe a decline in Australian impact, whereas the presented data (reprinted as Fig. 1) in our view do not support this? The original findings about the Australian funding policy introduced in the early 1990s did fit very well with intuition and expectations: if you ask for more output without demanding quality, you get more output without quality! These conclusions were quickly taken up in the science policy research community, and Australia was many times used as an example of a wrong policy, creating wrong incentives. Also Schneider, Aagaard and Bloch (2016) do so, and they make explicit that ‘The Norwegian system was specifically developed to counter adverse effects like those identified in the Australian case’. When something becomes part of the ‘accepted knowledge’, it strongly influences scientific observations, and that generally also holds in cases where the accepted knowledge is wrong.

We now have the interesting situation that a ‘publication counting system not taking quality into account’ performs better than ‘a system that does take quality into account’. How can we explain this? The criticism on the Australian model was rather intuitive: if

you ask for quantity without asking for quality, then you should get what you ask for, i.e. decreasing quality. But we may need to think more in terms of ‘counterintuitive effects’, which characterize social systems (Forrester 1995). One would intuitively think that attention for quality differences is positive for improving quality (the hope of the Norwegian system). But human beings are not good in ‘interpreting how social systems behave’, as ‘social systems belong to the class called multi-loop nonlinear feedback systems’ (ibid). How this exactly works in case of incentives for researchers need to be further investigated, and we suggest a possible hypothesis.

The Norwegian system measures not the quality of the paper, but the quality of the media it has been published in. Instead of looking at individual papers and their impact (quality), the journal is taken into account.³ As the impact of a journal and impact of a paper are different things, this approach is heavily disputed (Seglen 1994, 1997; Stephan, Veugelers and Wang 2017), and even proponents of the Norwegian model like Sivertsen (Zhang, Rousseau and Sivertsen 2017) agree with this. But it also may have behavioral effects that are counterproductive, as researchers may start to spend a lot of time and energy to adapt their papers to the ‘top’ journals. This is not necessarily ‘improving’ the paper, as it also requires adaptation to the ‘style’, and still many may fail to get it into such a ‘top’ journal. So the energy may be wasted, which better could have been used for more research—resulting in a better paper submitted to a ‘lower impact’ and often field specific journal which addresses often a more specialized audience. In the Norwegian case, we saw that the distribution of papers over the impact classes did not change. In the Australian case, the increased output was published in on average lower impact journals, but it had overall a larger impact, which supports our argument.

We conclude that the evidence in Schneider, Aagaard and Bloch (2016) shows that the Norwegian system stimulating output growth while taking ‘quality classes’ of this output into account performs less than the Australian system focusing only on output. Science policy makers may rethink what this means.

Notes

1. There is some discussion about when the policy was implemented (Van den Besselaar, Heyman and Sandström 2016, 2017), but we follow on this point here the same date as Schneider, Aagaard and Bloch 2016. But even if one would date the new policy 2 years earlier, the figure reveals that impact decline in the 1980s was stopped, and at the at the very least shortly after 1991 Australian impact started to rise.
2. These findings do confirm our own findings about the Australian case we reported elsewhere (Van den Besselaar, Heyman and Sandström 2016, 2017).
3. Or in case of books: the standing of the publisher.

References

- Forrester, J. W. (1995) ‘Counterintuitive Behavior of Social Systems,’ in: Forrester, J. W. (ed.) *World Dynamics*. Waltham MA: Pegasus Communications.
- Schneider, J. W., Aagaard, K., and 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: 244–56.
- Seglen, P. O. (1994) ‘Causal Relationship between Article Citedness and Journal Impact’, *Journal of the American Society for Information Science*, 45: 1–11.

- Seglen, P. O. (1997) 'Why the Impact Factor of Journals Should Not be Used for Evaluating Research', *BMJ*, 314: 497.
- Stephan, P., Veugelers, R., and Wang, J. (2017) 'Blinkered by Bibliometrics', *Nature*, 544: 411–2.
- Van den Besselaar, P., et al. (2016), 'Science Policy Through Stimulating Scholarly Output—Does it Work? Ismael Rafols. et al.,' in *Peripheries, Frontiers and Beyond; Proceedings STI 2016*, Valencia.
- Van den Besselaar, P., Heyman, U., and Sandström, U. (2017) 'Perverse Effects of Output-based Research Funding? Butler's Australian Case Revisited', *Journal of Informetrics*, 11: 905–918.
- Zhang, L., Rousseau, R., and Sivertsen, G. (2017) 'Science Deserves to be Judged by Its Contents, Not by Its Wrapping: Revisiting Seglen's Work on Journal Impact and Research Evaluation', *PLoS One*, 12/3: e0174205.