

available at www.sciencedirect.comjournal homepage: <http://www.elsevier.com/locate/scaman>

RESEARCH NOTE

Are more and better indicators the solution?

Margit Osterloh*, Bruno S. Frey

University of Zurich, Zurich, Switzerland

KEYWORDS

Scholarly evaluation;
Decision process;
Publishing

Summary We discuss Starbuck's proposal to improve decision processes in scholarly evaluation. While we agree that more variety is needed in evaluation committees, we suggest to enlist scholars from other research fields rather than people from outside academia. We disagree with the proposal that more and better indicators of research effectiveness will improve research. We argue that this even would lead to worse results than what is observed today. Attention would be deviated from the content of research, and intrinsic motivation, which is essential for good research, would be crowded out. We propose that evaluations that are based on indicators need to be pushed back. After a careful selection process, researchers need to be given the opportunity to pursue the research they consider to be fruitful.

© 2009 Elsevier Ltd. All rights reserved.

It is a pleasure for us to comment on William Starbuck's (2009) paper. It is a most competent, well-informed and relevant contribution and moreover – which is a rather rare occasion – fun to read. We strongly agree with his critique of how behavioural and social science is undertaken today. In particular, we think that his strictures against the “mass production of knowledge” and the “over-reliance on and misuse of statistical methods” is well taken. We address our comments to two suggestions in William Starbuck's concluding section “What to do?”.

First, William Starbuck proposes that researchers should “enlist people other than academics” in the evaluation of research to make researchers aware of the biases that infuse their work. We agree that different perspectives help to conduct more relevant research. There is indeed considerable evidence that a *variety* in opinions helps to avoid what has been called *groupthink* (Janis, 1982), i.e. the inability of evaluation committees that are composed of insiders to consider broader issues. To get a more open and relevant

view of what research has produced, persons from outside the narrow research field are required. However, it has long been established in the economics and sociology of science that an efficient governance of academia is characterized by a “republic of science” (Polanyi, 1962), which is self-organized by its peers, offers substantial discretion to scholars in choosing what they see as the most challenging scientific problems, and has a special reward structure called “taste for science” (Merton, 1973; Dasgupta & David, 1994). This reward structure is characterized by recognition by the peers and autonomy rather than by marketable outcomes. It is questionable whether people from outside academia share this “taste for science”. Of course, there are many examples in the history of science that peers have misjudged path-breaking research for a long time or have overestimated the value of research ideas. To mitigate such errors while enabling the necessary “view from outside” by people who share the “taste for science”, we propose to enlist people from inside academia, though from a different field of research.

Second, William Starbuck (2009) suggests that “The only way to overcome unreliable indicators of research effective-

* Corresponding author.

E-mail address: osterloh@iou.uzh.ch (M. Osterloh).

ness is to develop and disseminate reliable indicators". If this sentence is taken in a strict sense, it is difficult to object: reliable indicators are better than unreliable indicators. But if this opening statement of "What to do?" is taken more broadly – and we believe that this is Starbuck's intention – we disagree with respect to the evaluation of scholarly work. Even if indicators for research quality were perfect – which is not the case, since reviewers' biases, low inter-rate reliability and low prognostic quality are ubiquitous (Starbuck, 2005), and bibliometrics are "breathhtakingly naïve" (Adler, Ewing, & Taylor, 2008: 14) – evaluating research by indicators leads to even worse results than what is observed today. There are two main reasons.

The first reason is that a more intensive use of research indicators deviates attention even more from the *content* of research. What researchers are told is that it is essential to do well on the indicators – and that it does not really matter how and why. High scores in the measurement based on questionable data become the goal rather than a means of measuring quality. As a consequence, strategic reactions in the form of multiple tasking and counter-strategies of scholars and institutions are to be expected. Even if qualitative and quantitative measures worked perfectly, this problem could not be avoided.

The multiple tasking problem has been studied in economics extensively (Holmstrom & Milgrom, 1991): People maximize indicators that are easy to measure and disregard features that are hard to measure. An example is the "slicing strategy" when scholars divide their research results into a "least publishable unit" spreading them into as many papers as possible. This has been demonstrated in a study for Australia (Butler, 2003). The mid-1990s saw a linkage of the number of peer-reviewed publications to the funding of universities and individual scholars. The number of publications increased dramatically but the quality (measured by citations) decreased.

Counter-strategies go further than multiple tasking and are more difficult to observe. They consist in altering research behaviour itself in order to "beat the system", examples are:

- Scholars are induced to distort their results to please, or at least not to oppose, prospective referees. Frey (2003) calls this behaviour "academic prostitution".
- Reviewers are prone to judge papers more favourably that cite their own work approvingly and tend to reject papers threatening their previous work. Some editors pressure authors to cite their journals in order to raise their impact ranking (Monastersky, 2005).
- Creative and unorthodox research is discouraged and pedestrian research is encouraged. A referee process based on the opinions of average peers favours average research (Gillies, 2008). As a consequence, scholars are induced to produce predictable but unexciting results that will be accepted more easily by referees rather than path-breaking contributions.
- Homogenization of research endeavours takes place. For economics, Great Britain provides an example. The share of heterodox, not strictly neoclassical economics, has sunk drastically since the ranking of departments, based in part on their publication and citations, was established. The reason is that small and specialized journals are less

attractive for researchers due to their small impact factor (Lee, 2007).

The second reason why better indicators may lead to even worse research results is that the reliance on indicators leads to a systematic *crowding out of the intrinsic motivation to do research*. There is an extensive literature showing that intrinsic (rather than extrinsic) incentives are crucial for good research. This also corresponds to casual observations: it is difficult to think of a leading scientist who was not mainly, and sometimes entirely, motivated by his or her curiosity. Just consider the greatest natural scientists of the world, such as Newton or Einstein, or the greatest social scientists, such as Schumpeter or Keynes. The problem is that the intrinsic interest in doing research is crowded out when academics are evaluated based on extrinsic rewards that are contingent on indicators (Frey, 1997; Bénabou & Tirole, 2003).

For these two reasons, improving the quality of research indicators does not provide a useful solution to the basic problems of today's research. A more fundamental approach has to be considered (see more fully Osterloh & Frey, 2009): Evaluations by research indicators have to be pushed back, and researchers have to be given the opportunity to pursue the content and type of research they consider to be fruitful. Once young scientists have shown during a carefully conducted socialization and selection process that they are well qualified, creative and intrinsically motivated to do research, they should be given the opportunity to do the research they choose to undertake. Such kind of input control is not unusual in other professional fields that are characterized by the inappropriateness of objective output measures, like in legal institutions and governance agencies (e.g. Posner, *in press*). Though peer review has many shortcomings, it will still be necessary to screen the candidates for appointment and to decide on grant applications. To improve the decision processes, scholars from a different field of research should be enlisted in the evaluation committees.

It is clear that such a resource allocation may allow some academics to slow down their efforts. However – and this is what should count in academia – the rest of them produce useful and even path-breaking research, and they succumb less to faddishness.

References

- Adler, R., Ewing, J., & Taylor, P. (2008). Citation statistics, A report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics. Corrected version 6/12/08.
- Bénabou, R., & Tirole, J. (2003). Intrinsic and extrinsic motivation. *Review of Economic Studies*, 70, 489–520.
- 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.
- Dasgupta, P., & David, P. A. (1994). Towards a new economics of science. *Research Policy*, 23, 487–521.
- Frey, B. S. (1997). *Not just for the money*. Cheltenham, UK, Northampton, MA, USA: Edward Elgar.
- Frey, B. S. (2003). Publishing as prostitution? Choosing between one's own ideas and academic success. *Public Choice*, 116, 205–223.

- Gillies, D. (2008). *How should research be organised?* King's College London: College Publications.
- Holmstrom, B., & Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization*, 7, 24–52.
- Janis, I. L. (1982). *Groupthink: Psychological studies of policy decisions and fiascoes*. Boston: Houghton Mifflin.
- Lee, F. S. (2007). The research assessment exercise, the state and the dominance of mainstream economics in British Universities. *Cambridge Journal of Economics*, 31, 309–325.
- Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigation*. Chicago, IL: University of Chicago Press.
- Monastersky, R. (2005). The number that's devouring science. *Chronicle of Higher Education*, 52(8), A12.
- Osterloh, M., & Frey, B. S. (2009). Academic managerialism vs republic of science. Working Paper University of Zürich.
- Polanyi, M. (1962). The republic of science: Its political and economic theory. *Minerva*, 1(1), 54–73.
- Posner, R. A. (in press). From the new institutional economics to organization economics: With applications to corporate governance, governance agencies, and legal institutions. *Journal of Institutional Economics*.
- Starbuck, W. H. (2005). How much better are the most prestigious journals? The statistics of academic publication. *Organization Science*, 16, 180–200.
- Starbuck, W. H. (2009). The constant causes of never-ending faddishness in the behavioural and social sciences. *Scandinavian Journal of Management*, 25, 108–116.