



COMMENT:
AIDAN BYRNE

MEASURE FOR MEASURE

The system for assessing the link between science and innovation is flawed.

Few would dispute that the advancement of scientific knowledge leads to improvements in living standards. For some people, particularly those in politics or government, knowing that this link exists is just the beginning. They argue that to better engineer outcomes that benefit industry, governments or society, it is imperative to not just assert but precisely measure the interactions between science and innovation. Demonstrating this connection helps governments and institutions justify public expenditure on research.

In establishing a system for measuring the link between science and innovation, the underpinning rationale becomes vital. Unless the reason for measuring something is clearly articulated, an evaluation process is unlikely to provide any useful outcomes. Indeed, careless focus at the design stage can produce a perverse instrument that leads to undesirable practices, a phenomenon sometimes called Goodhart's law, in reference to British economist Charles Goodhart. Although British anthropologist, Marilyn Strathern, said it best: "When a measure becomes a target, it ceases to be a good measure".

PERVERSE INSTRUMENT

A case in point is the rise of university rankings over the last decade. Universities around the world pay close attention to lists that are, at best, determined by very crude and occasionally inappropriate measures of performance. Reporting these numbers in the form of rankings compounds the issue. It gives a false impression of precision, when the underlying data cannot reasonably distinguish the performance of institutions at such a granular level. For example, a ranking of 37 may look significantly better than a ranking of 66, but in most cases such a distinction does not reflect a real difference in the performance between two institutions.

To avoid this trap, the Australian government agency tasked with examining research performance, the Australian Research Council, uses ratings rather than rankings.

The evaluation of research excellence exercise (ERA) collects all academic outputs from the sector, including peer-reviewed journal papers, as well as books, reports and patents to provide ratings by discipline. It is unfortunate that when the data are released, they are quickly and misleadingly converted into rankings by others.

While gauging research excellence is a challenging task, it is relatively easy compared to quantifying the connection between research and innovation. Every discipline has its own accepted norms about what constitutes good research. No such shared understanding exists for connecting innovation and research. Efforts are further confused, and the subject of suspicion, when the motivation for the exercise is unclear.

**"WHEN A MEASURE
BECOMES A TARGET,
IT CEASES TO BE A
GOOD MEASURE."**

For example, one approach adopted by the UK in its system for assessing research quality, the Research Excellence Framework, examined case studies, in which institutions described the impact their research had on society. If the point is to show that research is capable of providing value for society, then examining case studies provides useful insights. But, if the goal is to provide indicators that help institutions refine and improve their approach to enhancing innovation, then a significant flaw emerges. Because it often takes decades to translate research into an activity that has an impact, case studies rely on academic and industry environments that

have probably changed, making them of limited use to inform future behaviour.

PROVIDING VALUE

Case studies are also flawed because they provide only a small sample of an institution's activity. Using them to allocate resources may reward the wrong people.

For an evaluation to be fair, and more than just an expensive PR exercise for research institutions, approaches need to involve the collection of more universal data that provide a holistic analysis. This concept underpins Australia's new approach to evaluate the impact of research, and the community's engagement with industry and the public.

This exercise, now in pilot phase, will collect a broad range of proxy indicators as measures of engagement, such as research income from industry sources. It will also examine a small number of case studies. Cumulatively, this information will provide evidence of the mechanisms that institutions have in place to enable scientific research to be translated into societally valuable products, services and amenities.

Although the pilot is in the early stages, it is clear that only a small range of indicators with any degree of reliability are available. This suggests that any ongoing assessment exercise needs to be adaptable, accommodating indicators that develop over time. Failure to do this will lock the exercise into relying on poor indicators.

Those creating evaluation instruments need to be mindful of how the research community is likely to respond to a measurement, and whether these responses create the desired outcome.

But, universities need to take some responsibility for how they use this information. They should be vigilant to avoid the trap of making decisions based on rankings, rather than addressing underlying issues. ■

Aidan Byrne is the provost at the University of Queensland, Australia, and the former head of the Australian Research Council.