

# The Value of Input Metrics for Assessing Fundamental Research

N. Tomecko <sup>a</sup> and D. Bilusich <sup>a</sup>

<sup>a</sup> *Joint and Operations Analysis Division, Defence Science and Technology Group, South Australia*  
Email: [daniel.bilusich@dst.defence.gov.au](mailto:daniel.bilusich@dst.defence.gov.au)

**Abstract:** Fundamental research is high pay-off, high-risk research aimed at game-changing discoveries. These research programs have long-term objectives and may not deliver their outcomes for many years. It has long been understood that small incremental gains in progress are often not a good indicator of the eventual achievement of a program's long-term objectives. So how can governments and corporations ensure that their investment is funding the right research programs?

There has been a significant amount of research into what research programs need to be successful. The findings of such research can be summarised in three points: programming (identifying clear research goals that only fundamental research can solve); ring-fencing (defining a contained program of work which is independent of other short term requirements, parallel income-generating consultancy or intermediate product development); and monitoring (evaluating the outputs and outcomes of research programs). The first two points are relatively easy to achieve, the third can be difficult. While research programs require monitoring to gauge which research groups are performing well and to inform future funding allocations. Output metrics (often called perverse metrics by frustrated researchers) such as the number of journal publications, the impact factors of published papers, or numbers of graduating PhD students, are typical measures used to assess research performance. However, when funding agencies publish the metrics they will use to determine funding allocations, researchers' need to remain competitive for funding results in a shift in their focus. Research focus shifts from meeting the primary objectives (research) to meeting short-term output requirements as dictated by the funding agency's metrics. Significant research shows the negative effect of these metrics on true, forward looking, fundamental research.

We argue that to stimulate fundamental research, performance metrics need to change their focus from monitoring outputs to monitoring inputs. The purpose of assessment is to fund research that is likely to achieve high level objectives. However, with fundamental research, past progress may not indicate future major discoveries. The timing of such discoveries cannot be controlled and therefore performance assessments should not focus on outputs. What can be controlled is the fundamental research-enabling environment; these inputs should be assessed to support future fundamental discoveries. An environment that fosters fundamental research will lead to fundamental research.

In this paper we present four input categories of measures relating to the research environment where assessments should focus. These categories cover various characteristics of the research environment that includes the following: the right people, adequate resourcing, defining the goals, allocating time, putting in place processes to encourage the research and promoting a culture of collaboration. For each of these categories we provide several examples of metrics that could be used.

**Keywords:** *Metrics, measures, assessing science, assessment of innovation, input metrics, research*

## 1. INTRODUCTION

Our progress as a society is closely tied to innovation and new discoveries. It is not surprising then, that governments as well as private organisations invest significant funding into activities that aim to discover new knowledge. The OECD spent an average 2.4% of their GDP on research in 2015 (OECD, 2017). However, there will always be a mismatch between the total funding available for research, and the demand for such funding from research organisations making assessments for funds allocation inevitable.

Funding bodies want to ensure that they receive the best return on their research investment. Generally, research programs are being assessed to ensure funding is given to organisations (and individuals) with the most highly rated research outputs. Funding allocation is reviewed regularly in line with these assessments in an attempt to ensure that high quality research is always achieved.

Another reason for assessing research is to ensure it is meeting its objectives. For example, Koier and Horlings (2015) found the objective of evaluating transdisciplinary research is to assess to what extent integration of knowledge was achieved (a research outcome measure). Measuring research outcomes is complex due to time lag and uncertain causality.

There are other reasons for research assessments not linked to research outputs or outcomes, namely, accountability. Research funded from public money needs to demonstrate that funding is being spent on resources directly for research. Commercial organisations are also accountable (for example to shareholders) and thus assess their research efforts.

For all of the above mentioned reasons, some assessment of research is required, but its benefits should outweigh the burden (Bird et al., 2005). Increasingly, evidence is emerging of the negative effects of regular assessments of research outputs (discussed in Section 2). Many studies discuss the negative impacts, but while there is a large body of research documenting these, there is little information on how to minimise them.

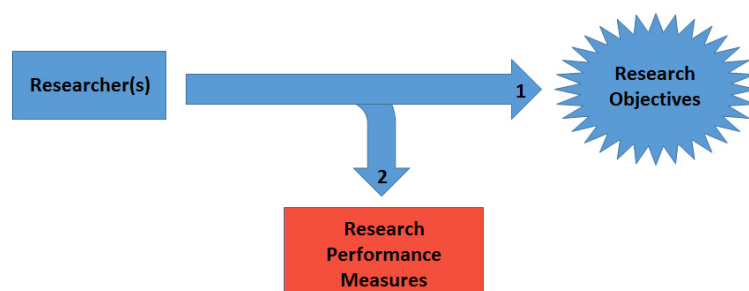
Some of this is understandable; the guide for best practice for assessment and metrics states there is no evaluation method that is universally applicable to all research (Australian Government, 2012). As such, most research relating to the assessment of research focuses on understanding the characteristics of a good evaluation scheme. This includes determining the frequency, the type of measures to be used, the weight that should be put on each metric and how to measure (e.g. peer assessment or bibliometrics). Many researchers advocate a multimethod approach; combining peer assessment and output metrics to evaluate research quality.

In this paper, we describe some issues with reporting/assessment, and discuss metrics to assess fundamental research which have minimal negative effects on the conduct of research. We advocate the inclusion of input measures in a multimethod evaluation approach.

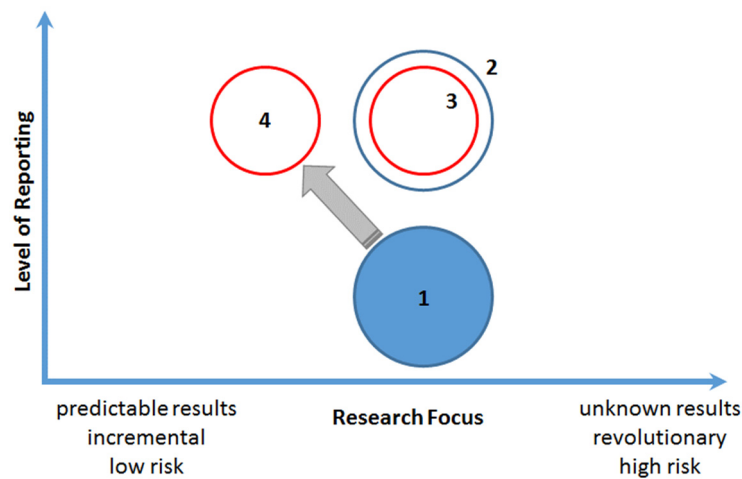
## 2. THE EFFECT OF REPORTING ON RESEARCH

There is often an assumption that the process of measurement does not influence the behaviour of individuals and institutions involved (Bird et al., 2005). However, there is strong evidence that contradicts this assumption, namely, that 'high stakes' performance assessment does indeed affect behaviour. Such side effects are often counter-productive (Bird et al., 2005). Behavioural change is a factor because no performance management scheme can be viewed in isolation from the incentives (designed or accidental) that exist alongside it (Bird et al., 2005).

Figure 1 is a representation of researchers working to achieve their research objectives. Where performance measures are put in place to assess the progress of this research (progress is represented by the arrow labelled '1'), particularly to determine which research groups to allocate future funding to, they can result in researchers modifying their behaviour to focus on meeting the performance metrics (arrow labelled '2'), diverting focus from their research objectives.



**Figure 1.** The effect of reporting and assessment on research activities.



**Figure 2.** The effect of reporting and assessment on research activities.

Figure 2 represents the impact of such measurement on the objectives. As an example, the blue circle labelled ‘1’ represents the level of research (the x-axis) a particular research group needs to focus on to achieve their high level research objectives. This is independent of any progress measurements placed upon them, and the area of the circle represents time allocated to conducting research. When a certain level of research performance measurement or progress reporting is implemented, the blue circle should simply move up the y-axis to the position of the blue ring labelled ‘2’ and the research should not change. However, this is not the case.

Firstly, reporting for the purposes of assessment such as preparing grant proposals, ensuring organisational databases are up to date, etc. is time consuming. A study in 2013 revealed that the cost of applying for €40 mil of annual grant money in the Netherlands was €9.5 mil, based on the salaries of those preparing the grant proposals (de Vrieze, 2017). In 2012, scientists in Australia spent 550 working years of researchers’ time preparing research grant proposals (of which only 20.5% were successful) (Herbert et al., 2013). Peer-reviewed assessment systems impose a high time investment (Butler, 2007). Therefore the time available to conduct the research is reduced, as represented in Figure 2 by the smaller red ring labelled ‘3’.

Performance measures can also influence behaviour and shift researchers’ focus to research that is more likely to meet the metrics used to assess research performance. This is represented in Figure 2 by the red ring labelled ‘4’; researchers shift towards research that guarantees results to meet performance monitoring measures (shifting left along the x-axis) at the expense of their original research objectives.

If measures to assess output, or progressive output, are employed and tied to funding, researchers face pressure to demonstrate outputs within a reporting period, or lose funding. The risk of losing funding therefore diverts research effort from more novel research in favour of research that more likely guarantees certain outputs, and thus guarantees continuation of funding. This generally results in incremental improvements and evolutionary, rather than revolutionary, outcomes.

To compound the problem, the pressure to demonstrate outputs can result in perverse behaviours as some researchers ‘game’ the system (Bird et al., 2005). Since the pressure is to *demonstrate* outcomes, not necessarily to actually *produce* them, these ‘perverse incentives’ result in unethical behaviour among researchers.

There are many examples of research performance schemes that have resulted in undesirable effects or behaviours within the research community. The “Evaluation methodology” (introduced by the Czech Republic in 2004 and evolved over subsequent years), a points-based system for awarding funding, where research outputs are weighted based on their perceived prestige, resulted in opportunistic behaviour by researchers (Good et al., 2015). A short-term increase in the number of publications (which is since slowing down) resulted from researchers publishing in lower impact journals in an effort to earn more points. A similar effect was observed in Australia, when a publication number-based system was introduced. While publication numbers increased, due to the shift to lower impact factor journals, Australia’s relative citation impact over the same period decreased (Good et al., 2015). According to Edwards and Roy (2017), and papers cited within, today’s research publications frequently suffer from lack of replicability because researchers ‘invent’ data points or results.

However, assessment and (regulatory) monitoring is necessary and can be helpful. For example, without some scrutiny, desired learning goals might not be achieved by Honours or PhD students. However, the equivalent assessment and monitoring on ambitious and novel exploratory research may be detrimental because many experiments may fail before an insightful and informative result that contributes to the development of knowledge is observed. The risk to the researcher is that there may be no direct progress made during a given performance measurement cycle, resulting in a negative assessment. Additionally, by its very definition, many ambitious (high-risk, high-payoff) programs will actually fail to achieve their objectives. To advance human

knowledge, only one out of several high-risk research programs needs to succeed. So long as one of several projects pushes the boundaries of knowledge by a significant amount, the entire investment in many such programs is worthwhile. It is therefore essential to determine the right type of reporting for each situation.

In Figure 2, fundamental exploratory research is on the right of the x-axis. The intended outcome of this type of, generally, curiosity-driven research is to develop new knowledge. Fundamental research can take significant time to transition from discovery to impact on society. In contrast, the main focus of commercial organisations is generally small innovations that are guaranteed to yield an improved product every season, or generate a new product launch cycle. This is essential to remain competitive in the marketplace. They may also conduct applied research that aims to advance the application of a specific technology which underpins their ability to launch new products in the future.

The spectrum of activities that incorporate some form of novelty is likely to be a continuum along the x-axis as illustrated in Figure 2. For this paper, we adopted the categories in Table 1 to span this spectrum. Each category has a different aim. Therefore, while broadly following the pattern described in Figure 2, assessment/reporting and the associated metrics will have a different impact on its objectives. For each of those categories, measures should obviate rather than create, perverse behaviour (Bird et al., 2005). Of these, fundamental and seminal research are particularly affected by assessment/monitoring, because previous research progress is often not a good indicator of future research outcomes.

**Table 1.** Research, Innovation and Development activity types.

Activity type	Activity output	Activity outcome	Example
Product improvement	Better products	Incremental performance improvement	iPhone 4, 4s
Product development	New products	Innovation (new uses of existing knowledge)	First iPhone (smart phone)
Applied research	New applications	Solving problems using existing techniques (possibly from different areas)	Touchscreens for phones, or flickpage feature on screen
Fundamental research	New knowledge	Describing and discovering new facts/features about the physical world	Miniaturise “current in glass” concept
Seminal research	New field of knowledge	Defining a new perspective with which to observe the physical world	Conductive glass

### 3. A SOLUTION – ENABLING THE RIGHT ENVIRONMENT

Assessing fundamental research requires measures that minimise the burden on, and perverse behaviours by, researchers, and supports research to achieve planned objectives, i.e. staying on the right hand side of the x-axis in Figure 2. By definition, fundamental research programs have long-term objectives where the full outcomes of potentially years of research may not be evident for a significant period. It has long been understood that small incremental gains in progress are often not an indicator of reaching the primary objectives (Australian Government, 2012).

The intent of current assessment processes for research and innovation is to attempt to guarantee future research outcomes by rewarding researchers with high outputs. Refer back to Figure 1 where the ‘research performance measures’ box is placed across the arrow labelled ‘1’. However, this method of assessment is flawed. So how can an assessment scheme foster ‘good’ future outcomes? To address this, we need to step back to first principles of what contributes to the likelihood of future outcomes.

We cannot control when a researcher will produce a novel result (the researcher themselves cannot do that), yet that is what we are trying to measure. Instead, we need to identify the areas that we can control that contribute to research success and direct our assessments to those. The elements of the research system that we can control are the inputs into the research, and the environment in which research is conducted. In Figure 1, this would be represented by the measures box shifting to the left beside the ‘researcher’ box.

While research success cannot be guaranteed, an environment can be provided that gives the best possible opportunity for generating the intended research outcomes. This means we shift focus from measuring research outputs, to measuring the environment in which research is conducted. These are input measures. Our assessment needs to reward researchers/organisations that invest in these inputs.

What are the input metrics? What environments have been linked to the desired research outcomes? What are the input measures that contribute to success? The ingredients define the environment: for example, the skills, the equipment and resources available to researchers, their ability to transfer knowledge between each other in an environment free from competition, the behaviours and decisions that define how we encourage and nurture research and innovation such as conference funding and opportunities for collaboration.

While evidence is emerging to support the proposition that improving certain inputs yields better research outcomes, there has not yet been an explicit systemic shift from measuring outputs to measuring inputs. Based on literature, the use of input metrics falls into three categories: 1) papers that describe input metrics without using the term 'input', 2) papers that discuss the measurement of inputs, but fail to make a link to research outputs, and 3) papers that discuss the importance of inputs, but do not advocate measuring them explicitly.

As an example of the first category, in Portugal in the early 2000s, there was a goal to create critical mass (the input) in each scientific area by bringing together large research consortia through networks across institutions. As a result, the number of papers in Portugal rose from 3792 (in 2000) to 10081 (in 2010) (Ramos and Sarrico, 2016). However, they do not refer to critical mass as an input.

In the second category, in Finland, output based metrics are only 5% of formula-based funding. There are two input indicators included in the funding allocation: number of researchers and researcher mobility. Similarly, in Norway, input metrics form part of the evaluation (Good et al., 2015). In Spain, researcher collaboration is measured via a number of co-authorship index metrics (Diaz-Faes et al., 2016). While differences between disciplines are observed regarding the amount of international collaboration in particular, discussion of linking the metric to research assessment is limited to cautions of any interpretations being cognisant of disciplinary differences. The European review of frontier (fundamental) research concluded that special pull factors were necessary to foster fundamental research. *Ring-fencing* fundamental research, i.e. enabling researchers to dedicate time to research, unencumbered by competing objectives, was the most important factor identified (Leijten et al., 2009).

For the third category, Adams et al. (2006) reviewed possible metrics for measuring innovation, and included inputs as one category. However, they proceed to cite frameworks that considered some measures from each of these categories independently rather than suggesting that they could be integrated into a more effective evaluation framework. Others have described the use of inputs as part of performance measures conceptually rather than describing an evaluation framework (Janssen et al., 2011).

We define input metrics as those that focus on the fundamental inputs to research, and measure the environment within which research is conducted. These metrics fall into the following categories:

- **Metrics related to the researchers.** These relate to the ability of the people charged with research to do the research. Do the researchers have the right skills? Do they have (are we giving them the opportunity to form) the right connections? Do they collaborate with others within and beyond their organisations and with the users of their research?
- **Metrics related to the facilities.** These metrics relate to sufficient access to equipment and facilities necessary for the research.
- **Time allocation to research.** These metrics relate to researchers having the ability to dedicate time to research, where other commitments (such as teaching, client work, administration) do not compete with their research efforts.
- **Process.** Culture, Innovation process, activity: metrics that measure the opportunity for research and innovation.

Examples of various metrics in each of those categories are shown in Table 2. The shift from output metrics to input metrics also implies a shift in the interpretation. The use of output metrics invokes the paradigm of achieving certain thresholds, while the use of input metrics invokes the paradigm of looking for evidence.

**Table 2.** Research, Innovation and Development activity types and metric applicability

Metric Category	Sub-category	Metric Examples
People	Critical Mass of researchers	Number of research and technical staff Number of PhD students currently enrolled Ongoing recruitment activities Visiting researcher numbers
	Collaboration indicators	Conference attendance Conduct of joint activities (e.g. workshops and seminars) Number of different fields represented in the activity Shared resources (such as high grade equipment)
	Ongoing relationship with other researchers	Participation in communities of practice Informal networks
Facilities		Adequate/appropriate equipment Future facility plans
Time allocation to research		Existence of strategic plans for research Long term appointment of researchers Opportunity for sabbaticals Removal from other major commitments and other income generating activities
Process	Culture	Opportunity and encouragement to collaborate (e.g. existence of collaborative research projects) Opportunity to attend conferences and seminars
	Process	Existence of defined pathways to realise novel ideas Processes to convert solutions/discoveries into tangibles

We aim to reduce perverse behaviours – so are input metrics helping us do that?

There already is a tacit understanding that creating the right environment leads to facilitating fundamental research outcomes without diverting focus to incremental results. Such philosophy was used to found the Santa Fe Institute (Waldrop, 1992) and is the philosophy behind Centres of Excellence. Additionally, evidence is starting to emerge that for cognitive work, traditional motivators do not work (Ariely et al., 2009), which is part of the reason many output-based incentive schemes are counterproductive. For cognitive work, people are motivated by intrinsic goals – they want to be part of something bigger and have a purpose – if their need for a purpose is met, they will be motivated to immerse themselves in the work. If we create the right environment and let scientists focus on what they do best, research outcomes will follow.

We acknowledge that the solution to the problem is not as simple as focusing on input metrics. What is needed is a multimethod approach to research assessment; we advocate that input metrics can contribute significantly to a multimethod approach because they remove some of the burden and add value to the assessment process.

#### 4. CONCLUSIONS

Assessment of research is an unavoidable requirement. The focus should shift from the current approach of primarily output-based assessment to a multimethod approach that includes input measures. To guarantee new discoveries, the right environment for fundamental research needs to be created through the provision of the right people, adequate resourcing, defining the goals, allocating time, putting in place the processes to encourage the research and promoting a culture of collaboration. It is the input measures that can be controlled and refined/optimised to furnish the best conditions for fundamental research to flourish and generate valued outcomes.

Inputs play a major role in achieving research objectives, and should therefore be included as part of a multimethod approach for funding allocation deliberations.

## ACKNOWLEDGMENTS

We acknowledge Alexandra Shekhter, David Blunt, Nicola Broderick, Vinh Dang, Joseph McConnell and Lachlan Carlile for inspiring this paper.

## REFERENCES

- Adams, R., Bessant, J. and Phelps, R. (2006). Innovation management measurement: A review. *International Journal of Management Reviews*, 8, 21-47.
- Ariely, D., Gneezy, U., Loewenstein, G. and Mazar, N. (2009). Large stakes and big mistakes. *The Review of Economic Studies*, 76, 451-469.
- Australian Government, (2012). Best Practice Guide to Evaluation of Science and Innovation Initiatives, Department of Industry, Innovation and Science. Available at: <http://industry.gov.au/science/councilsandcommittees/Pages/Library%20Card/CCIBestPracticeGuidetoEvaluationofScienceandInnovationInitiatives.aspx>
- Bird, S.M., Cox, D., Farewell, V.T., Goldstein, H., Holt, T. and Smith, P.C. (2005). Performance indicators: good, bad, and ugly. *Journal of the Royal Statistical Society A*, 168, 1-27.
- Butler, L. (2007). Assessing university research: a plea for a balanced approach, *Science and Public Policy*, 34, 565-574.
- Diaz-Faes, A., Bordons, M. and van Leeuwen, T.N. (2016). Integrating metrics to measure research performance in social sciences and humanities: the case of the Spanish CSIC. *Research Evaluation*, 25, 451-460.
- Edwards, M.A. and Roy, S. (2017). Academic research in the 21<sup>st</sup> century: maintaining scientific integrity in a climate of perverse incentives and hypercompetition. *Environmental Engineering Science*, 34, 51-61.
- Good, B., Vermeulen, N., Tiefenthaler, B. and Arnold, E. (2015). Counting quality? The Czech performance-based research funding system. *Research Evaluation*, 24, 91-105.
- Janssen, S., Moeller, K. and Schlaefke, M. (2011). Using performance measures conceptually in innovation control. *Journal of Management Control*, 22, 107-128.
- Herbert, D.L., Barnett, A.G. and Graves, N. (2013). Funding: Australia's grant system wastes time. *Nature*, 495, 314.
- Koier, E. and Horlings, E. (2015). How accurately does output reflect the nature and design of transdisciplinary research programmes? *Research Evaluation*, 24, 37-50.
- Leijten, J., Roseboom, H. and Hofer, R. (2009). More Frontier Research For Europe, A Venture Approach for Funding High Risk – High Gain Research, Report SIS 2009 ERABSTUDY, European Commission, DG Research.
- OECD (2017). Gross domestic spending on R&D (indicator). Available at: <http://data.oecd.org/rd/gross-domestic-spending-on-r-d.htm>
- Ramos, A. and Sarrico, C.S. (2016). Past performance does not guarantee future results: lessons from the evaluation of research units in Portugal. *Research Evaluation*, 25, 94-106.
- de Vrieze, J. (2017). With this new system, scientists never have to write a grant application again. *Science*. 13 April 2017. Available at: <http://www.sciencemag.org/news/2017/04/new-system-scientists-never-have-write-grant-application-again>
- Waldrop, M.M. (1992). *Complexity. The Emerging Science at the Edge of Order and Chaos*. Simon & Schuster, New York, NY.