Universiteit Leiden
The Netherlands

Effects of Indicators on Knowledge Production in Astronomy

Heuritsch, J.

Citation
Heuritsch, J. (2018). Effects of Indicators on Knowledge Production in Astronomy. *Sti 2018 Conference Proceedings*, 101-106. Retrieved from https://hdl.handle.net/1887/65237

Version: Not Applicable (or Unknown)
License: Leiden University Non-exclusive license
Downloaded from: https://hdl.handle.net/1887/65237

Note: To cite this publication please use the final published version (if applicable).

STI 2018 Leiden

23rd International Conference on Science and Technology Indicators

"Science, Technology and Innovation Indicators in Transition"

STI 2018 Conference Proceedings

Proceedings of the 23rd International Conference on Science and Technology Indicators

All papers published in this conference proceedings have been peer reviewed through a peer review process administered by the proceedings Editors. Reviews were conducted by expert referees to the professional and scientific standards expected of a conference proceedings.

Chair of the Conference

Paul Wouters

Scientific Editors

Rodrigo Costas
Thomas Franssen
Alfredo Yegros-Yegros

Layout

Andrea Reyes Elizondo
Suze van der Luijt-Jansen

The articles of this collection can be accessed at https://hdl.handle.net/1887/64521

ISBN: 978-90-9031204-0

© of the text: the authors
© 2018 Centre for Science and Technology Studies (CWTS), Leiden University, The Netherlands

This ARTICLE is licensed under a Creative Commons Atribution-NonCommercial-NonDetivates 4.0 International Licensed

Effects of Indicators on Knowledge Production in Astronomy [1]

Julia Heuritsch

heuritsch@dzhw.eu
German Centre of Higher Education Research and Science Studies (DZHW), Schützenstraße 6A, 10117 Berlin; Humboldt Universität zu Berlin, Research Group “Reflexive Metrics”, Institut für Sozialwissenschaften, Unter den Linden 6, 10099 Berlin

Abstract

Nine interviews were conducted with astronomers from Leiden University, and a document analysis was performed on relevant institutional (self-) evaluation documents, annual reports, and CVs of the interviewees. The aim was to perform a qualitative study about the relationship between the research behaviour of astronomers and how their science is being evaluated. This study encompassed the funding and publication system as well as the indicators used to measure the scientific output, its quality and the research performance. It sheds light on how astronomers define high-quality research and how they think that creating knowledge of value is encouraged or hampered by the evaluation processes. The research shows that astronomers are realists who define scientific quality on the basis of “truth” and are driven by curiosity. These two factors make up their intrinsic values and motivation to perform Astronomy. Publication pressure, arising from the requirements of “the system”, creates an extrinsic motivation to perform. This results in premature publications, low readability and replicability, risk aversion and a focus on quantity rather than quality. Hence, indicators do not merely represent quality, but also co-constitute what counts as good research. While such constitutive effects of indicator use on research behaviour and content are observed, there is no indication that the astronomer's intrinsic values are co-constituted. This gives rise to a discrepancy between what is being measured by indicators and what astronomers define as scientific quality; the so-called "evaluation gap". Findings on constitutive effects and the evaluation gap in Astronomy lays out the conceptual groundwork for further empirical research and for policy advice on alternative evaluation practices and innovative indicators with the aim of bridging the "evaluation gap".

Introduction

The sociology of science found that policies have constitutive effects on how science is performed (Peter Dahler-Larsen, 2014). In other words, indicators and rewards introduced by policies shape the process of knowledge production is produced. It has been argued that there may be a discrepancy between what is being measured by indicators and the quality of the

[1] Acknowledgements: I would like to thank the interviewees for their willingness to participate in this study and for dedicating research time that is even more valuable in times of publication pressure. I would also like to thank Sarah de Rijcke (CWTS, Leiden), who supervised this project.

scientific content as perceived by the researchers in the field; the so-called evaluation gap (Wouters, 2017). In order to meet the targets set by indicators, scientific quality may be sacrificed. Goal displacement, gaming, information overload, questionable authorship practices, unhealthy competition and aversion to risky/ innovative projects are only a few examples of possible consequences (Rushforth & De Rijcke, 2015; Laudel & Gläser, 2014).

Astronomy is the perfect science to study from a meta-perspective. It is dedicated to basic research, strongly instrumentalized, involves large collaborations, and the use of (open) archives and huge datasets. Astronomy asks highly fundamental questions which inspire both scientists and the public at large. What effects do incentives due to funding, publication practices & indicators have on the knowledge production processes and research behaviour in Astronomy? What effects do they have on research quality?

This project was designed to find the answer to that question. When we know how policies with their indicators and reward systems can affect quality, we are one step closer to know how to steer research. This article summarises[2] the results while the complete report can be found in ArXiv[3]. Direct quotes of the interviewees will be given between double quotation marks.

Methods

This research consists of interviews and a document analysis. Semi-structured interviews were conducted with (international) researchers (4 faculty members, 2 postdocs, 1 PhD) and 2 Master students from Leiden Observatory (Sterrewacht). The Master programme at the Sterrewacht is very research intensive, requiring the student to write two Master theses in total, which is the reason why they are also interesting subjects for this study. In order to shed light on the evaluation gap, questions were developed such that an astronomer's definition of quality versus what is measured by indicators can be studied. Topics include career steps, project funding, exposure to assessments, research evaluation, the publication and funding system, different stages of the knowledge production process – from planning, via doing the research to publishing – and the meaning of quality. Each topic was introduced by one overarching question, followed by several potential follow-up questions.

The participating researchers were invited via email and all names are anonymized. All interviews, 80-100 minutes in length, were fully transcribed into electronic form, summarised and coded) according to Grounded Theory. These codes represent themes which emerged by combining sensitivity towards existing literature on constitutive effects of indicator use with insights from the data.

The interview data were complemented with a document analysis of materials collected online or made available via the informants, including CVs of the interviewed researchers, annual reports (1998 to 2015), (self-) evaluation reports of the Dutch Astronomy institutes and their umbrella organisation NOVA.

[2] Content of this paper was first published in the Spring 2018 edition of 'ROOM – The Space Journal' (15 #1). For further details see: www.room.eu.com
[3] https://arxiv.org/abs/1801.08033

What is scientific quality for an astronomer?

In order to understand the extent of a potential evaluation gap and how indicators shape knowledge production in Astronomy, we must investigate the intrinsic values and motivation of an astronomer and compare this with what is required by the evaluation system. Only if we know what research quality means for an astronomer, we can investigate whether indicators have constitutive effects on quality.

The study found that astronomers generally conduct science for the sake of curiosity and "pushing knowledge forward", that is searching for the truth and discovering structures of nature. Astronomers are realists, who assume a reality independent from the observer, arising from (physical) causal laws. As Astronomy is the study of the universe and its building blocks, it seeks to answer the most 'fundamental' questions to set a basis for 'truth of everything'.

> "But that moment of – you know – mystery, that is a scientific experience in the sense that there is only one thing that you accept in that moment, that's the *truth*, you want to know the *real answer*. And no excuses, only the real answer matters. And that is what *drives science*; we only want to know the real answer."

The notion of truth and the quest to push knowledge forward both result from the astronomers' curiosity to understand the universe. It follows that high quality in research means that there is a correspondence between reality and the scientific theory. For a realist, truth and scientific quality are objective and scientific integrity is implied.

> "I think in terms of what constitutes *good science* and what is *academic integrity*, all those things don't change – they are pretty close to *absolute values* I would say."

Thus the study found that the definition of high quality research is based on three criteria:

1. Asking an important question for the sake of understanding the universe better and to push knowledge forward.

2. Using clear, verifiable and sound methodology.

3. Clear communication of the results in order for the community to make use of them.

As obvious that definition may seem to an astronomer, scientific quality is hard to measure, and numbers are easy to look at (Benedictus & Miedema, 2016). That is why indicators serve as proxies to evaluate scientific quality and performance. In order to compare this evaluation with astronomers' intrinsic motivation and values, the study investigated what the funding, publication and career systems value and which indicators are used, according to the astronomers.

The Evaluation Gap in Astronomy

It turned out that the so-called Matthew effect prevails in those systems. This effect, the "Chicken-or-egg problem", means that past output determines future success. Output is measured in terms of publication- & citation rates and impact factors, and acquired grants and observation time. Sexy topics, or "the fashion of the day" are often favoured over important basic research. To survive the climb up the career ladder, an astronomer has to acquire recognition through grants (includes observation time) and publish enough first author papers. Ranking of the universities of previous job positions influence further career development. The Matthew effect leads to a "Golden Child Trajectory", where the 'ideal' career in Astronomy is a straightforward climb of the tenure track. This often involves committing to a professional life in the hamster wheel of the "cycle of observing, analysis and publishing".

From this investigation of values two opposite notions of science emerge. The first is the 'ideal' image of science, where astronomers are driven by their curiosity and the search for truth, limited only by epistemic restrictions such as technical possibilities. The other notion is the image of a 'system' of science, comprising the funding, publication and evaluation system, with values that are not in line with the astronomer's intrinsic values. One reason for the discrepancy between the 'ideal' and the 'system' notions of science is that scientific discovery is naturally unpredictable, where a scientists may end up with no desired outcome. Effort cannot be monitored efficiently, so assessment cannot be based upon it. Therefore, a scientist is rewarded and funded for quantitative achievement instead (Rosenberg & Nelson, 1994). The discrepancy between those two notions of science gives rise to the evaluation gap.

Constitutive effects of indicators on knowledge production in Astronomy

The presence of the evaluation gap has shaping consequences on the research behaviour and knowledge production in Astronomy, so-called constitutive effects. Dahler-Larsen (2014) distinguishes five main categories of constitutive effects of indicators: indicators define interpretive frames and world views *(A),*social relations and identities *(B),* content *(C),* time frames *(D)* and change in their meaning as a consequence of their use *(E).*

The remainder of this article will portray constitutive effects arising from the evaluation gap in Astronomy and relate them to one or more of those categories.

According to the interviewees in the study, output in the form of papers and its quality assessment through quantitative indicators such as publication & citation rates, defines the value of an astronomer *(B).* The need for this kind of output to survive on the career ladder ("publish-or-perish"), however, causes publication pressure *(D),* which may have psychological effects on the researcher *(B)* and constitutive effects on the (quality of the) content *(C).* The latter may include cutting up publications in order to publish more ("salami slicing"), premature publishing, and non-replicable papers *(C).* In most cases, according to the interviewees, those effects have a negative effect on research quality. Publication pressure however, can also have positive effects, focusing and confining the research question. Salami slicing can be beneficial for good communication and readability of research results. The interviewees, however, remarked that Astronomy, like Psychology, finds itself in a replication-crisis. This is because of the lack of incentives to publish information needed for replication, such as code used for analysis. Prematurely published information makes output even less readable and reproducible, reducing content quality.

Quantitative indicators define the landscape of success and its inverse: the landscape of failure *(A).* The interviewees have a hard time defining 'failed research', due to the very risky nature of research. They are only confident to describe what bad research is – the opposite of good quality research according to their definition. In contrast, the community and the system do have a definition of failed research, viz. the opposite of successful research as measured by indicators. The use of indicators then causes a shift in what counts as new discoveries in the community, from ‘anything new’ to ‘publishable results’. Hence, as long as negative results (e.g. non-detections) can’t be put in a context (“tailoring”) where they become publishable, they are regarded as worthless: the research project failed and the researcher feels like a failure *(A, B)*. This is despite the fact that in many fields in Astronomy non-detections are far more common than detections. Because those are hardly made public, astronomers express their frustration with the ‘wheel being reinvented’ and hence resources wasted. Especially young researchers can’t afford to take on too risky research projects, which causes risk aversion (“playing it safe”) and a tendency to prefer sexy topics to equally important non-sexy ones. This again has effects on research agendas and content *(C)*.

Thus the evaluation system undermines astronomers’ values, shifting the focus from high quality, robust, replicable and well-communicated research results to an information overload of premature results on sexy topics, with conclusions that are less robust, replicable, and transparent than aspired. As a consequence, an astronomer’s motivation also shifts to an output orientation where safe and accessible projects become the driver. While “the publication is not the aim – [it] is a means to showing what your methodology is” it does become an aim. Producing high quality research “to know and understand better” and communicating this knowledge to the community is what makes up an astronomer’s intrinsic motivation to conduct research. However, the need to survive the climb up the career ladder gives an extrinsic motivation to perform research, which is oriented towards hitting the required targets *(B, E)*.

> “And ahh … if [publishing] was not so important [to keep your standing] … I mean I would still publish my papers [but] it *gives a different motivation* to it, right? As a scientist you just want to publish your papers, because you are a scientist and you think this is important for science: ‘This is the result, this is what defines the process of science’. […] Yeah, I have my doubts about the usefulness of that system.”

However, the study also makes clear that, while indicators give an extrinsic motivation to an astronomer to perform, their constitutive effects do not reach as far as to affect an astronomer’s intrinsic values to a noteworthy extent. As a consequence, a third notion of science arises: coping with the system. Astronomers always try to manage a balancing act between their intrinsic values and the requirements of the system.

The conclusion can be drawn that Goodhart’s law holds true: “when a measure becomes a target, it ceases to be a good measure” *(E)*. Future policies on indicators, rewards and incentives must take into account that evaluation leads to an evaluation gap and has constitutive effects on research quality. Stakeholders and policy makers must take into consideration that quality can hardly be measured in quantified terms. Bibliometric data should therefore be complemented by alternative ways to find high quality research(ers). In addition, current and future studies (will) investigate alternative indicators and ways to evaluate research.

References

Dahler-Larsen, P. (2014), “Constitutive Effects of Performance Indicators: Getting beyond unintended consequences”, *Public Management Review,* 16:7, p.969-986

Rushforth, A.D. & De Rijcke, S. (2015). “Accounting for Impact? The Journal Impact Factor and the Making of Biomedical Research in the Netherlands”, *Minerva 53*, p.117-139

Laudel, G. & Gläser, J. (2014), “Beyond breakthrough research: Epistemic properties of research and their consequences for research funding”, *Research Policy 43,* 1204-1216.

Wouters, P. (2017), “Bridging the Evaluation Gap”, *Engaging Science, Technology, and Society 3*: p.108-118

Benedictus, R., & Miedema, F. (2016), “Fewer Numbers, Better Science”, *Nature,* Vol 538, p.453-455

Rosenberg, N. & Nelson, R. (1994), “American Universities and technical advance in industry”, *Research Policy 32*: p.323-348