Unveiling the ecosystem of science: Towards an integrative valuation model of the many roles of scientists

Nicolas Robinson-Garcia^{a,*}, Rodrigo Costas^b, Thed N. van Leeuwen^b, Tina Nane^a

^a Applied Mathematics (DIAM), TU Delft, Delft, Netherlands
^b CWTS, Leiden University, Leiden, Netherlands

Abstract

There is increasing evidence on the misuse and abuse of quantitative indicators in the current scientific reward system. Alternatively, more qualitative approaches, use of case studies or the design of indicators more sensitive to societal and scientific needs have been suggested. In this article we analyze the bases of such criticisms and motivations for changing the reward system by focusing on the assessment of individuals. We explore alternative models proposed or in use and identify common characteristics. Based on this we propose a valuation model by which we can systematically organize and prioritize performative aspects of scientists and consider other factors which may affect or might relevant for research policy. We finally test our model in a series of case studies based on six academic units.

1. Introduction

The ecosystem of science is defined as a system with interconnected entities integrated in a larger social system with which a bidirectional influence is exerted (Fortunato et al. 2018). Scientists, as drivers of the research enterprise, are immersed in an increasingly diverse set of tasks and activities with the purpose of producing, communicating and transferring knowledge to society. This involves a series of skills which go beyond their intellectual or technical capabilities (e.g., negotiation skills, communication skills, social engagement). The introduction of national and supranational research evaluation systems (Hicks 2012) carved the path to introducing measures of assessment in the hope that they would help monitor and maximize such research efforts. While peer review has traditionally been the main mechanism of quality control (Bornmann 2011), there has been an increasing use of bibliometric indicators. According to Gläser et al. (2002), this is due to the fact that,

 ${\it Email~address:}~{\tt N.Robinson@tudelft.nl}~({\tt Nicolas~Robinson-Garcia})$

^{*}Corresponding author

"[b]y applying quantitative performance indicators, actors in science can be compared according to their performance more or less independent of peers' judgments, and science policy can reach conclusions seemingly independent of the scientific community to which an actor belongs" (pp. 29).

An overemphasis on quantitative and specifically, bibliometric indicators, has led to a situation in which academic careers are assessed and rewarded relying solely on individuals' capacity to publish highly cited research articles in highly ranked journals (McKiernan et al. 2019; Wouters 2014). This unidimensional view of the research enterprise has been harshely criticized (DORA 2014; Hicks et al. 2015; Wilsdon et al. 2015), arguing mostly but not exclusively, on the methodological and conceptual limitations they have when assessing individuals. But there are more profound and serious effects derived from current research evaluation schemes.

First, the focus on publications may lead to a task reduction, where scientists neglect those types of work which are not recognized (Rijcke et al. 2016). As consequence, there is a goal misplacement as the researcher's efforts go into gaining higher scores in the measures instead of pursuing their own research agenda. Second, quantitative indicators are not always adequately used and introduced in research evaluation schemes (Ràfols et al. 2016), which means that seemingly 'egalitarian' criteria can greatly hamper researchers from certain fields or specific profiles (Gläser and Laudel 2007). Also, because such indicators can be gamed (Fister, Fister, and Perc 2016), counteracting rules to avoid misbehaviours can be even more harming (Robinson-Garcia and Amat 2018). Third the focus on publishing in journals with a high impact factor in tenure and promotion (McKiernan et al. 2019) leads to systematic biases on the type of scientific knowledge produced. Rafols et al. (2012) demonstrated systematic suppressing of interdisciplinary research when using journal-based indicators for research evaluation. Piñeiro and Hicks (2015) showed that, in the case of Spanish Sociology, policies promoting publication in high impact journals indexed in the Web of Science would penalize locally-oriented research. Chavarro, Tang, and Ràfols (2017) reported that non-mainstream journals, which play an important role in national science systems, are negatively affected by these policies.

Finally, because a single set of criteria is defined for all scholars, research evaluation schemes tend to promote a notion of the 'excellent scientist' who is capable of outperforming in all facets (Olmos-Peñuela, Benneworth, and Castro-Martínez 2016), threatening the diversity of scientific profiles researchers needed in the science ecosystem (Gläser et al. 2002; Woolley and Robinson-Garcia 2017).

This article proposes an alternative model for evaluating scientists' performance. This approach balances between a conceptually-informed framework and a methodologically viable operationalization. We first provide evidences on the specific aspects that are being negatively affected by research evaluation schemes by reviewing previous empirical work. After reviewing the different proposals made in the literature, we propose a set of five performative dimensions which

consider as policy-relevant when conducting research assessment of individuals. A key aspect of the model has to do with the networked nature of scientific work and hence, the heterogeneity of research profiles developed. Hence, how research trajectories develop and understanding how specialization on specific roles might affect such categories is a key point for our model. Also external factors affecting individuals' performance along with effects of evaluation on personal features such as gender, nationality or age, are considered as they can be directly or indirectly affected by research policies and evaluation schemes. The model is implemented in a multiple-case study of six organisational subunits (i.e., departments, laboratories) from two Dutch universities. This work contributes to existing research on the scientific workforce and research careers in an inefficient evaluative context (Hammarfelt and Rushforth 2017) which threatens diversity of scientific profiles, and hence of outputs and impacts (Olmos-Peñuela, Benneworth, and Castro-Martínez 2016; Woolley and Robinson-Garcia 2017).

2. Pervasive effects attributed to the scientific reward system

Bibliometric indicators were incorporated in research evaluation exercises in the early 1970s (De Bellis 2014). Based on the notion of citation and productivity as predictors of recognition and academic success (Reskin 1977; Merton 1968), their use soon expanded to from the United States to Europe through the United Kingdom's first national Research Assessment Exercises in the 1980s (Luukkonen 2002). Citation and publication-based indicators were seen as an objective and measurable way in which to assess performativity and impact. Merton's normative framework of recognition built the grounds for the use of citations and proxies for research quality. But it ignored the effects the use of such indicators would have on scientists behavior and consequently, on the production of novel societally relevant scientific knowledge (Benedictus, Miedema, and Ferguson 2016). Furthermore, scientific communication is drastically transforming due to the emergence of the Open Access movement and the development of social media and research infrastructure (i.e., data banks, repositories). This has also affected the business model behind the publishing industry (e.g., author fees). Hence, it is no longer feasible nor advisable to assess qualitatively research outputs nor rely on journal or journal-derived metrics solely (Ioannidis and Khoury 2014).

Effects on the scientific workforce

The first and most obvious unintended effect is the emergence of the so-called *Publish or Perish* culture. Dalen and Henkens (2012) indicate that scientists working in high pressure environments perceive an external pressure to publish, which makes them pay excessive attention only to "tasks that benefit or glorify the individual (publishing internationally, being cited by other scholars)" instead of those "tasks that benefit larger groups" (pp. 1291). This leads to a task reduction (Rijcke et al. 2016) that negatively affect other tasks expected from scholars such as teaching, knowledge transfer (Lewison 2005) or engagement with societal stakeholders (Dalen and Henkens 2012; Robinson-Garcia, Leeuwen, and Rafols 2018).

Second, funding opportunities affect how research is conducted (Luukkonen and Thomas 2016). Academic freedom is mediated by the expectations the community has based on what it considers as scientific merit (Polanyi 2000). As evaluation schemes shape such expectations, scientists' goals can deviate from what is considered more relevant or interesting, to what pays off best. By using journal rankings or rewarding high citation rates, evaluation schemes pressure scientists to modify their research agendas as well as their strategies to maximize their efforts. For instance, an emphasis on novelty is a contributing factor for the lack of replication studies, under-reporting of false positives, or testing of prior published results (Nosek, Spies, and Motyl 2012). Also, strategic publishing can lead to select topics which are not necessarily societally relevant but that become scientific hypes, hence becoming very profitable in terms of citations (Rousseau, García-Zorita, and Sanz-Casado 2013).

Third, the notion of performativity becomes entangled with the indicators used to measure it. In this case, publication output and citation impact. A suspicion comes then from anything that deviates from the norm, not bein able to interpret correctly those cases, as "there may be favoritism in some competitions and smart hiring in others" (Abramo, D'Angelo, and Rosati 2015, 772). Furthermore, these policies tend to emphasize and promote the *superstar* or leading excellent scientist (Olmos-Peñuela, Benneworth, and Castro-Martínez 2016), ignoring the increasing need for team players, also referred to as the 'middle author' (Mongeon et al. 2017). As competition instead of collaboration is promoted, those middle authors end up being excluded from the academic career (Milojević, Radicchi, and Walsh 2018).

Effects on knowledge production

Critique	Evidences	Reference
Misalignment between societal needs and research	Desktop research and expert panel workshop	Moher et al., 2018

1. Research quality: Low quality, overproduction, lack of replication, research integrity, fraud, hypercompetition (less collaboration),

http://print-cfa.dk/wp-content/uploads/2019/06/ncrm-manchester-2019-jws.pdf

2. Transparency. P-hacking, reproducibility, under-reporting of negative results

Authors like Moher et al. (2018) or Nosek (2015) refer to research integrity, and in the case of the latter, refer to conflict of interest of researchers as they might be interested on publication even if results are not accurate. This would lead to Ioannidis' famous paper on the reproducibility crisis and false results.

3. Societal challenges: Not aligned with societal needs but with greater impact within scientific community.

4. Long-term agendas. Publish or perish culture is short-sighted.

3. Towards a valuation model on the many roles of scientists

First include review on literature Examples of dimensions can be given from the book by Bastow, Tinkler & Dunleavy (2014)

Bozeman, Dietz & Gaughan (2001) present their evaluation model at the individual level as follows:

- Internal resources: cognitive skills (this can be criticized as it is impossible to operationalize as they present it i.e., ability to synthesize), S&T knowledge and context skills (learnt through experience). The authors indicate that each resource may have n dimensions.
- S&T Capital: Which has to do with the network of scientists and the intrinsic value they have as well as the role the individual plays in such network
- S&T Human Capital and Life Cycles: The authors recognize that researchers' profile will be different at different stages of their career.
- Corley et al (2017) introduce a cultural dimension to the model which refers to variables such as gender, nationality, race, discipline or socio-economic status.

ACUMEN presents a portfolio, a kind of CV format designed for assessing individual performance. It combines qualitative and quantitative information. It offers space for a narrative where an individual tells her own story. It then distinguishes between three aspects of an academic's career: expertise (methods, areas of theory, etc.), outputs (publications, patents, etc.) and impacts (citations, awards, etc.). Furthermore, it includes an individual's age. For these three aspects it includes quantitative indicators. The perspective here is that individuals report their CV and the ACUMEN portfolio basically structures and offers recommendations on how this should be provided. Other than that it does not give clear indications on how the assessment should be performed, it seems to rely on experts' judgment.

Look into Whitley

Look into Evaluative Inquiry

Career trajectories Stephan & Levin -> In terms of productivity peaks and decline

Laudel, G., Bielick, J., & Glà ¤ser, J. (2018). 'Ultimately the question always is: "What do I have to do to do it right?" 'Scripts as explanatory factors of career decisions. Human Relations, 0018726718786550. https://doi.org/10.1177/0018726718786550

Valuation models of scientific activity

I have changed the name of the section to valuation so that it alignes with the ISSI poster. This means that the variables considered in the table should also be modified and reflect on the aspects valued by each of these models.

Interesting the model presented by Boyer (1990) and used in the paper by Herman & Nicholas (2019) published in El profesional de la informaciÃ³n. Furthermore, should consider the term *facet* instead of referring to *dimensions*, it might be more accurate.

Evaluation model	Output/outcome	Unit of analysis	References
Bibliometric assessment Economic analysis	Production and impact	Scalable	
Knowledge value alliances Research portfolios	Journal articles?	Communities of scholars (e.g., labs)	Bozeman & Rogers Wallace & Rafols, 2015
Productive interactions	Process-oriented	Research projects	Spaape & van Drooge, 2011
S&T Human Capital		Scalable	Bozeman, Dietz & Gaughan, 2001

Evaluative dimensions

Scientific engagement

Collaboration ties and diversity (number of unique collaborators and intensity), position in network, (co-authorship, acknowledgment, etc.), strength of tie?, interdisciplinarity, application of research, member of committees, reviewer, etc.

$Social\ engagement$

Non-academic collaboration ties (either through publications or not i.e., reference patterns), some altmetric indicators (policy briefs but also maybe twitter classes? Or potentially Ed's ABC score), spin-offs,

Capacity-building

Productivity, leadership, independence, scientific impact, funding

Trajectory/Background?

Geographic mobility, cognitive mobility, sectoral mobility (These last three could be seen not as attributes to the researcher but their capacity to work with diverse people), career status

Research practices (Cross-sectoral)

OA publications, data sharing, outreach (e.g., The Conversation, blogging, tweeting), publication of working papers, proceedings, mentoring

Some references

Alperin, J. P., Nieves, C. M., Schimanski, L., Fischman, G. E., Niles, M. T., & McKiernan, E. C. (2018). How significant are the public dimensions of faculty work in review, promotion, and tenure documents? Recuperado de https://hcommons.org/deposits/item/hc:21015/

July 3rd, researchers, 2018|Early career, education, H., science, O., evaluation, R., & Comment, R. policy|1. (2018, julio 3). Making research evaluation processes in Europe more transparent. Recuperado 20 de febrero de 2019, de https://blogs.lse.ac.uk/impactofsocialsciences/2018/07/03/making-research-evaluation-processes-in-europe-more-transparent/

Schimanski, L. A., & Alperin, J. P. (2018). The evaluation of scholar-ship in academic promotion and tenure processes: Past, present, and future. F1000Research, 7, 1605. https://doi.org/10.12688/f1000research.16493.1

Confounding effects

Luukkonen, T., & Thomas, D. A. (2016). The 'Negotiated Space' of University Researchers' Pursuit of a Research Agenda. Minerva, 54(1), 99-127. https://doi.org/10.1007/s11024-016-9291-z

Include here motivations of scientists -> puzzle, ribbon, gold and prosocial motivations.

Personal features

Order dimensions from time independent to dependent. Features within dimensions

Paper by Balen et al 2012 Also paper by Smith and Nosek 2015

4. Research design

Sample

To test the reliability of the model, we selected a technical university and a general university. Furthermore, these subunits belong to three different fields (Medical Sciences, Engineering Sciences and Social Sciences), a

Data sources

Analyses

5. Results

Subsections are tentative

Profiling of scientists

Field specificities

Career trajectories

Table with number of scholars in each lab by source (e.g., with Google Scholar profile) and by status.

- Scatterplots combining publications from different sources
- $\bullet\,$ Showcase dimensions for research teams or labs -> Internal organization and division of labor
- Showcase dimensions life cycles -> Different stages in career
- Attribution Plos One -> Scientific engagement
- Differences on profiles by fields.
- Agencies

6. Discussion

Tentative topics to discuss

• Diversity of roles or different career stages?

References

Abramo, Giovanni, Ciriaco Andrea D'Angelo, and Francesco Rosati. 2015. "The Determinants of Academic Career Advancement: Evidence from Italy." *Science and Public Policy* 42 (6): 761–74. https://doi.org/10.1093/scipol/scu086.

Benedictus, Rinze, Frank Miedema, and Mark W. J. Ferguson. 2016. "Fewer Numbers, Better Science." *Nature* 538 (7626): 453–55. https://doi.org/10.1038/538453a.

Bornmann, Lutz. 2011. "Scientific Peer Review." Annual Review of Information Science and Technology 45 (1): 197–245. http://onlinelibrary.wiley.com/doi/10.1002/aris.2011.1440450112/full.

Chavarro, Diego, Puay Tang, and Ismael Ràfols. 2017. "Why Researchers Publish in Non-Mainstream Journals: Training, Knowledge Bridging, and Gap Filling." Research Policy 46 (9): 1666–80. https://doi.org/10.1016/j.respol.2017. 08.002.

Dalen, Hendrik P. van, and Kène Henkens. 2012. "Intended and Unintended Consequences of a Publish-or-Perish Culture: A Worldwide Survey." *Journal of the American Society for Information Science and Technology* 63 (7): 1282–93. https://doi.org/10.1002/asi.22636.

De Bellis, Nicola. 2014. "History and Evolution of (Biblio)Metrics." In Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly

Impact, edited by Blaise Cronin and Cassidy R. Sugimoto, 23–44. Cambridge, MA: MIT Press.

DORA. 2014. "San Francisco Declaration on Research Assessment." https://sfdora.org/.

Fister, Iztok Jr, Iztok Fister, and Matjaž Perc. 2016. "Toward the Discovery of Citation Cartels in Citation Networks." *Frontiers in Physics* 4. https://doi.org/10.3389/fphy.2016.00049.

Fortunato, Santo, Carl T. Bergstrom, Katy Börner, James A. Evans, Dirk Helbing, Stasa Milojevic, Alexander M. Petersen, et al. 2018. "Science of Science." *Science* 359 (6379): eaao0185. https://doi.org/10.1126/science.aao0185.

Gläser, Jochen, and Grit Laudel. 2007. "Evaluation Without Evaluators." In *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems*, edited by Richard Whitley and Jochen Gläser, 127–51. Sociology of the Sciences Yearbook. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-1-4020-6746-4_6.

Gläser, Jochen, Grit Laudel, Sybille Hinze, and Linda Butler. 2002. "Impact of Evaluation-Based Funding on the Production of Scientific Knowledge: What to Worry About, and How to Find Out." Expertise for the German Ministry for Education and Research 31: 2002.

Hammarfelt, Björn, and Alexander D. Rushforth. 2017. "Indicators as Judgment Devices: An Empirical Study of Citizen Bibliometrics in Research Evaluation." Research Evaluation 26 (3): 169–80. https://doi.org/10.1093/reseval/rvx018.

Hicks, Diana. 2012. "Performance-Based University Research Funding Systems." Research Policy 41 (2): 251–61. http://www.sciencedirect.com/science/article/pii/S0048733311001752.

Hicks, Diana, Paul Wouters, Ludo Waltman, Sarah de Rijcke, and Ismael Rafols. 2015. "The Leiden Manifesto for Research Metrics." *Nature* 520 (7548): 429–31. https://doi.org/10.1038/520429a.

Ioannidis, John P. A., and Muin J. Khoury. 2014. "Assessing Value in Biomedical Research: The PQRST of Appraisal and Reward." *JAMA* 312 (5): 483–84. https://doi.org/10.1001/jama.2014.6932.

Lewison, Grant. 2005. "Beyond SCI Citations – New Ways to Evaluate Research." *Current Science* 89 (9): 1524–30. https://www.jstor.org/stable/24110919.

Luukkonen, Terttu. 2002. "Research Evaluation in Europe: State of the Art." Research Evaluation 11 (2): 81–84. https://doi.org/10.3152/147154402781776871.

Luukkonen, Terttu, and Duncan A. Thomas. 2016. "The 'Negotiated Space' of University Researchers' Pursuit of a Research Agenda." *Minerva* 54 (1): 99–127. https://doi.org/10.1007/s11024-016-9291-z.

McKiernan, Erin C., Lesley A. Schimanski, Carol Muñoz Nieves, Lisa Matthias, Meredith T. Niles, and Juan Pablo Alperin. 2019. "Use of the Journal Impact Factor in Academic Review, Promotion, and Tenure Evaluations." e27638v2. PeerJ Inc. https://doi.org/10.7287/peerj.preprints.27638v2.

Merton, Robert K. 1968. "The Matthew Effect in Science the Reward and Communication Systems of Science Are Considered." *Science* 159 (3810): 56–63.

https://doi.org/10.1126/science.159.3810.56.

Milojević, Staša, Filippo Radicchi, and John P. Walsh. 2018. "Changing Demographics of Scientific Careers: The Rise of the Temporary Workforce." *Proceedings of the National Academy of Sciences* 115 (50): 12616–23. https://doi.org/10.1073/pnas.1800478115.

Mongeon, Philippe, Elise Smith, Bruno Joyal, and Vincent Larivière. 2017. "The Rise of the Middle Author: Investigating Collaboration and Division of Labor in Biomedical Research Using Partial Alphabetical Authorship." *PLOS ONE* 12 (9): e0184601. https://doi.org/10.1371/journal.pone.0184601.

Nosek, Brian A., Jeffrey R. Spies, and Matt Motyl. 2012. "Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth over Publishability." *Perspectives on Psychological Science* 7 (6): 615–31. https://doi.org/10.1177/1745691612459058.

Olmos-Peñuela, Julia, Paul Benneworth, and Elena Castro-Martínez. 2016. "Does It Take Two to Tango? Factors Related to the Ease of Societal Uptake of Scientific Knowledge." *Science and Public Policy* 43 (6): 751–62. https://doi.org/10.1093/scipol/scw016.

Piñeiro, Carla López, and Diana Hicks. 2015. "Reception of Spanish Sociology by Domestic and Foreign Audiences Differs and Has Consequences for Evaluation." Research Evaluation 24 (1): 78–89. http://rev.oxfordjournals.org/content/24/1/78.short.

Polanyi, Michael. 2000. "The Republic of Science: Its Political and Economic Theory." Minerva~38~(1): 1–21.

Rafols, Ismael, Loet Leydesdorff, Alice O'Hare, Paul Nightingale, and Andy Stirling. 2012. "How Journal Rankings Can Suppress Interdisciplinary Research: A Comparison Between Innovation Studies and Business & Management." Research Policy, Exploring the Emerging Knowledge Base of 'The Knowledge Society', 41 (7): 1262–82. https://doi.org/10.1016/j.respol.2012.03.015.

Ràfols, Ismael, Jordi Molas-Gallart, Diego Andrés Chavarro, and Nicolas Robinson-Garcia. 2016. "On the Dominance of Quantitative Evaluation in 'Peripheral' Countries: Auditing Research with Technologies of Distance." SSRN Scholarly Paper ID 2818335. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=2818335.

Reskin, Barbara F. 1977. "Scientific Productivity and the Reward Structure of Science." *American Sociological Review* 42 (3): 491–504. https://doi.org/10.2307/2094753.

Rijcke, Sarah de, Paul F. Wouters, Alex D. Rushforth, Thomas P. Franssen, and Björn Hammarfelt. 2016. "Evaluation Practices and Effects of Indicator Use—a Literature Review." Research Evaluation 25 (2): 161–69. https://doi.org/10.1093/reseval/rvv038.

Robinson-Garcia, Nicolas, and Carlos B. Amat. 2018. "¿Tiene Sentido Limitar La Coautoría Científica? No Existe Inflación de Autores En Ciencias Sociales Y Educación En España." Revista Española de Documentación Científica 41 (2): 201. https://doi.org/10.3989/redc.2018.2.1499.

Robinson-Garcia, Nicolas, Thed N. van Leeuwen, and Ismael Rafols. 2018. "Using Almetrics for Contextualised Mapping of Societal Impact: From Hits to

Networks." Science and Public Policy 45 (6): 815–26. https://doi.org/10.1093/scipol/scy024.

Rousseau, Ronald, Carlos García-Zorita, and Elias Sanz-Casado. 2013. "The H-Bubble." *Journal of Informetrics* 7 (2): 294–300. https://doi.org/10.1016/j.joi.2012.11.012.

Wilsdon, J., Liz Allen, Eleonora Belfiore, Philip Campbell, Stephen Curry, Steven Hill, Richard Jones, et al. 2015. "The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management." HEFCE. http://doi.org/10.13140/RG.2.1.4929.1363.

Woolley, Richard, and Nicolas Robinson-Garcia. 2017. "The 2014 REF Results Show Only a Very Weak Relationship Between Excellence in Research and Achieving Societal Impact." *Impact of Social Sciences Blog.* https://blogs.lse.ac.uk/impactofsocialsciences/2017/07/19/what-do-the-2014-ref-results-tell-us-about-the-relationship-between-excellent-research-and-societal-impact/.

Wouters, Paul. 2014. "The Citation: From Culture to Infrastructure." In Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact, edited by Blaise Cronin and Cassidy R. Sugimoto, 47–66. Cambridge, MA: MIT Press.