https://doi.org/10.51408/issi2025_091

The Effects of Research Evaluation: Do Researchers' Perceptions Align with Evidence?

Giovanni Abramo,[1] Ciriaco Andrea D'Angelo,[2] Emanuela Reale,[3] Antonio Zinilli[4]

[1]*giovanni.abramo@unimercatorum.it*
Universitas Mercatorum, Laboratory for Studies in Research Evaluation
Piazza Mattei 10, 00186 Roma (Italy)

[2]*dangelo@dii.uniroma2.it*
University of Rome "Tor Vergata", Dept of Engineering and Management
Via del Politecnico 1, 00133 Roma (Italy)

[3]*emanuela.reale@cnr.it*, [4]*antonio.zinilli@cnr.it*
CNR-IRCrES, Consiglio Nazionale delle Ricerche, Research Institute on Sustainable Economic Growth,
Via dei Taurini 19, Roma (Italy)

Abstract

This study examines the alignment between researchers' perceptions of the Italian Scientific Habilitation (ASN) and the bibliometric evidence regarding its impact on scientific productivity in STEMM disciplines. The ASN, introduced in 2012, serves as a key evaluation tool for academic promotions in Italy, aiming to enhance research productivity and quality, and contrast favoritism. Employing a mixed-methods approach, we compare survey data from academics with bibliometric analyses of publication output over two five-year periods (2008–2012 and 2013–2017). The findings reveal significant misalignments: while bibliometric evidence indicates measurable productivity increases following the introduction of the ASN, survey responses suggest that many researchers perceive little to no impact.

The divergence between perception and evidence varies across demographic and disciplinary contexts. Younger researchers and early-career academics report stronger perceived and measurable productivity increases, reflecting their reliance on the ASN for career progression. In contrast, older researchers show measurable gains in bibliometric analyses but often do not attribute these improvements to the evaluation system. Disciplinary differences also emerge: fields such as Medicine and Engineering exhibit high productivity gains in both perception and evidence, while disciplines like Physics and Mathematics demonstrate significant bibliometric increases but low perceived impact.

This mismatch carries critical implications for research evaluation practices. For researchers, it highlights a potential erosion of trust in evaluation systems, particularly among those who feel their contributions are undervalued. This discontent may lead to disengagement or counterproductive behaviors, such as prioritizing short-term outputs or engaging in unethical practices like self-citation, citation networks, or salami-slicing publications. For institutions, the findings underscore the need to tailor evaluation practices to accommodate disciplinary differences and to recognize diverse contributions beyond publications, such as teaching and societal impact.

At the policy level, the study advocates for a more inclusive and transparent evaluation framework. Recommendations include integrating qualitative assessments with bibliometric metrics, developing discipline-specific evaluation criteria, and addressing disparities in gender, geography, and institutional resources. Efforts to enhance transparency and communication in evaluation systems could bridge the gap between perception and evidence, fostering greater trust and legitimacy.

Despite its contributions, the study has limitations. The survey data captures subjective perceptions that may be influenced by personal biases, while bibliometric analyses rely on productivity proxies

that overlook qualitative aspects of research. Future research should employ longitudinal and qualitative methods to explore the underlying causes of misalignment and its impact on academic behavior.

By addressing the roots of the mismatch between perception and evidence, this study provides actionable insights for designing evaluation systems that align with academic values, promote equity, and incentivize long-term innovation.

Introduction

Research evaluation tools have become indispensable for assessing the pursuit of research policy goals and strategic objectives. They focus mainly on key dimensions of research performance such as productivity, quality, and impact of academic work (de Diego et al., 2024). These systems influence various decisions, create specific individual incentives, and stimulate organisational and management changes. However, their implementation has sparked an ongoing debate about their unintended consequences and the extent to which they align with the broader goals of scientific inquiry (de Rijcke et al., 2016). Central to this debate is the question of whether researchers' perceptions of these systems match empirical evidence regarding their effects. Misalignments between perception and evidence can distort academic priorities, undermine equity, and inhibit the cultivation of diverse intellectual landscapes.

In this article, we explore the complex dynamics between perception and evidence in the context of research evaluation. In particular, we intend to contrast outcomes related to the changes in research productivity (increase or decrease of productivity) from a survey-based study with those arising from bibliometric pictures, taking Italy as a field of observation since the country was recently interested in the heavy introduction of research assessment. In this work, by research productivity, we mean the publications produced by a researcher over a given time period, as this is the most widely accepted definition in academia and, therefore, suitable for use in a survey.[1] In particular, we concentrate on a research evaluation exercise named the Italian Scientific Habilitation (ASN).[2] Introduced for the first time in 2012, it enables habilitated individuals to be selected for positions of Associate professors and Full professors in Italian universities. Therefore, the evaluation exercise analysed in this paper is strongly related to the academic career. Our investigation focuses exclusively on STEMM fields, which have distinct publication practices and research evaluation dynamics compared to other disciplines. Above all, they are particularly well-suited for bibliometric evaluation.

The development and proliferation of research evaluation metrics have transformed academic ecosystems. Metrics such as the journal impact factor (Garfield, 1972), citation counts (Bornmann & Daniel, 2008), and the h-index (Hirsch, 2005) were initially designed to complement qualitative assessments of research quality. According to a few scholars, their widespread adoption has led to an over-reliance on quantitative measures, often reducing complex scholarly contributions to narrow,

---

[1] For a more detailed definition of research productivity, we refer the reader to Abramo and D’Angelo (2014).

[2] https://abilitazione.mur.gov.it/public/index.php?lang=eng.

one-dimensional scores. This “metric fixation” (Muller, 2018) has contributed to several well-documented issues, including the reinforcement of existing inequalities, a bias toward mainstream disciplines, and the undervaluation of less measurable dimensions of academic work, such as teaching and mentorship (McKiernan et al., 2016). Other scholars hold that the problem with metrics is that they are applied by individuals without professional expertise, while evaluative scientometricians know well in which circumstances to adopt scientometrics and in which to recur to other methods (Abramo, 2024; Ioannidis & Maniadis, 2023).

Researchers’ perceptions of these evaluation systems often reflect frustration with their perceived rigidity, bias, and opacity. Surveys indicate that many researchers feel pressured to prioritize short-term outputs, such as publishing in high-impact journals, over long-term goals, such as fostering innovation or addressing societal challenges (Nicholas et al., 2017; Fire & Guestrin, 2019). Moreover, qualitative studies suggest that evaluation systems can create misaligned incentives, encouraging practices such as salami-slicing publications or favoring “safe” research over more exploratory or interdisciplinary work (Sahel, 2011; Brembs et al., 2013). While these perceptions are widely reported, empirical evidence presents a more nuanced picture of the effects of evaluation systems, highlighting both their benefits and drawbacks (Abramo & D’Angelo, 2021; Seeber et al., 2019).

Empirical studies reveal that evaluation metrics can effectively identify high-impact research and facilitate comparisons across disciplines and institutions (Waltman, 2016). However, they also underscore significant limitations. For instance, citation-based metrics are heavily influenced by field-specific publication practices, with some disciplines inherently generating fewer citations than others (Moed, 2005). Additionally, gender and geographic disparities persist, with women and researchers from the Global South often receiving less recognition and fewer citations, even when their work is of comparable quality (Larivière et al., 2013). These findings challenge the assumption that using metrics for research evaluation is neutral or universally applicable, suggesting that researchers' perceptions of bias may be well-founded.

The mismatch between perception and evidence in research evaluation has profound implications. When researchers perceive evaluation systems as unfair or misaligned with academic values, it can erode trust, reduce motivation, and lead to gaming behaviors that undermine the integrity of the scientific process (Smaldino & McElreath, 2016). Conversely, efforts to address this misalignment—such as initiatives promoting responsible research assessment (DORA, 2012; Hicks et al., 2015) and the use of narrative CVs (Moher et al., 2022)—have shown promise in fostering more equitable and holistic evaluation practices, which may mitigate these negative effects.

This paper aims to examine the misalignment between the perceptions of researchers and empirical evidence in research evaluation, specifically in relation to the impact of the ASN on scientific productivity, understood as the increase in scientific publications since its introduction.

To achieve this, within the context delineated above, the paper addresses the following question: “Is there a misalignment between researchers' perceptions and empirical evidence regarding the effects of research evaluation on productivity, when controlling for individual and contextual factors?"

The findings of the study can help formulate actionable strategies for bridging the gap between perception and evidence. By integrating insights from bibliometric research, sociology of science, and policy studies, we aim to provide a comprehensive understanding of how research evaluation systems shape academic behavior. Addressing the misalignment between perception and evidence is not merely a matter of improving metrics or processes; it is essential for restoring trust, promoting inclusivity, and ensuring that research evaluation serves its intended purpose of advancing knowledge and societal well-being. Through this lens, we aim to contribute to the ongoing dialogue on building evaluation systems that align with the values and realities of the research community.

The paper is organized as follows. In the next section, we will illustrate the methodological issues of the two proposed analyses and, in the following, the main results of the analyses and their comparison. The concluding section summarizes the main findings and illustrates the authors' considerations about implications and future developments.

Methods

This paper uses both survey and bibliometric analyses to explore the factors influencing the impact of the ASN on scientific productivity in STEMM disciplines. The analyses share a consistent framework of independent variables, ensuring comparability between the subjective perceptions captured in the survey and the outcomes derived from bibliometric data. Using the same set of independent variables, we aim to provide a comprehensive understanding of how individual, institutional, geographic, and disciplinary factors shape the perception of the ASN and its measurable effects.

The factors or independent variables included in the analyses are:

- Individual factors: Gender (male vs female) and age groups (<35, 35–44, 45–54, 55–65, and >65, with the oldest group serving as the reference category).
- Institutional size: universities are categorized as large- (reference category), medium-sized, and small-sized.
- Geographic location: Regions are categorized as North, Centre, and South (reference category).
- Disciplinary areas: The analysis includes 10 (STEMM)[3] of the 14 Italian university disciplinary areas (CUN), with Physics (CUN 2) serving as the baseline category.[4] We exclude from the analysis the areas of social sciences

[3] 1 - Mathematics and computer science, 2 - Physics, 3 - Chemistry, 4 - Earth sciences, 5 - Biology, 6 - Medicine, 7 – Agricultural and veterinary sciences, 8 - Civil engineering, 9 - Industrial and information engineering, 10 - Psychology.

[4] Physics is chosen as the baseline category for two reasons: i) Physics is a well-established field with relatively standardized research and publication practices. It provides a consistent benchmark for comparison with other disciplines that may have more diverse or variable practices; ii) Physics is known for its high volume of publications and collaborations, often within large international research

and arts and humanities, due to the limited coverage in bibliographic repertories of the research output in these areas (Mongeon & Paul-Hus, 2016; Archambault et al., 2006). For the area "Historical, Philosophical, Educational, and Psychological Sciences," only the subarea of Psychology is included in the analysis, as eligible for bibliometric analysis.

In the survey analysis, we assess the perceptions of researchers on whether the ASN has influenced their scientific productivity. Respondents were asked to consider the last ten years of their career, which means from date back until 2012 when ASN was introduced in Italy. In contrast, the bibliometric analysis measures actual changes in productivity, using a binary outcome variable indicating whether there was an increase in publication output between two five-year periods (2008–2012 and 2013–2017), i.e. after the introduction of the ASN. By combining these approaches, we can compare the perceptions with evidence, identifying both areas of alignment and divergence.

The survey

The data for this study were collected through a national survey conducted in Italy between 2020 and 2021. The survey used a structured questionnaire administered to a probabilistic sample of academics from Italian universities in the disciplinary areas under observation. The survey collected information on the effects of the ASN, focusing on individual adaptation or response, as well as respondent characteristics (e.g., gender, age, academic position) and institutional contexts (e.g., university size). For geographic distribution, the adopted classification is into three main macro-areas: North, Centre, and South. Regarding academic ranks, the study included five positions introduced by the Gelmini Law (L. 240/2010): researcher, type A researcher (RTD-A), type B researcher (RTD-B), associate professor, and full professor. In the following Table 1, a detailed breakdown of the survey dataset is provided.

Table 1. Breakdown of the dataset (822 professors) by personal and contextual variables.

Variable	Level	Share
Gender	F	36.1%
	M	63.9%
Age	Less than 35	1.7%
	35-44	19.6%
	45-54	34.1%
	55-65	34.5%
	Over 65	10.1%
Univ. size	Big	47.5%
	Medium	33.8%
	Small	18.7%
Univ. location	South	26.3%
	Center	26.8%
	North	46.9%

teams. Its citation practices and publishing norms are relatively well-aligned with bibliometric indicators commonly used in evaluation systems like the ASN.

We applied a logit model to analyze the likelihood of response and to identify the factors influencing respondents' perceptions of the ASN's impact on their productivity. This approach allows us to derive a regression equation capable of predicting the category each academic falls into, based on the explanatory variables. The dependent variable in this study was constructed using the survey question: "*In the past ten years, to what extent have the following factors influenced the quantity of your publications?*"

This question captures a range of influences on scientific productivity, including, but not limited to, the ASN. Factors considered include, for instance, the need to align with ASN requirements, gaining a competitive edge in securing research funding, participating in national or international research projects, and increasing academic visibility. These additional factors provide a comprehensive view of the various motivations and external pressures that may impact the quantity of publications.

To isolate the effect of the ASN from other factors, we focused specifically on respondents who reported an increase in productivity and explicitly attributed this change to the ASN. By narrowing the analysis to this subgroup, we were able to disentangle the impact of the ASN from other influences, allowing for a more targeted assessment of its role in shaping research output. This approach ensures that our findings reflect the specific contribution of the ASN, separate from broader or overlapping factors. The analysis focuses on 822 respondents belonging to STEMM scientific areas.

The bibliometric analysis

Our dataset comprises 26,217 professors (assistant, associate, or full) from Italian universities, who held tenured positions in STEMM fields continuously from 2008 to 2017. Table 2 shows their distribution by academic field and rank, based on data as of December 31, 2012, i.e. around the time the ASN was introduced for the first time in Italian academia. Table 3 summarizes the relative frequencies of personal variables (gender and age) and contextual variables (size and location of the university of affiliation).

Table 2. Dataset of the bibliometric analysis. Breakdown by field and academic rank.

Field*	Assistant prof.	Associate prof.	Full prof.	Total
1 – MATH	926 (38.5%)	794 (33.0%)	687 (28.5%)	2407 (9.2%)
2 – PHYS	620 (38.5%)	623 (38.6%)	369 (22.9%)	1612 (6.1%)
3 – CHEM	1024 (45.9%)	763 (34.2%)	446 (20.0%)	2233 (8.5%)
4 – EARTH	354 (44.6%)	286 (36.1%)	153 (19.3%)	793 (3.0%)
5 – BIOL	1725 (48.0%)	1070 (29.8%)	796 (22.2%)	3591 (13.7%)
6 – MED	3514 (49.0%)	2193 (30.6%)	1461 (20.4%)	7168 (27.3%)
7 – AGRVET	1034 (42.9%)	788 (32.7%)	587 (24.4%)	2409 (9.2%)
8 – CIVENG	435 (36.6%)	428 (36.0%)	326 (27.4%)	1189 (4.5%)
9 – INDENG	1422 (35.9%)	1377 (34.7%)	1167 (29.4%)	3966 (15.1%)
11 – PSYCH	353 (41.6%)	271 (31.9%)	225 (26.5%)	849 (3.2%)
Total	11407 (43.5%)	8593 (32.8%)	6217 (23.7%)	26217

* *1-Mathematics and computer science, 2-Physics, 3-Chemistry, 4-Earth sciences, 5-Biology, 6-Medicine, 7-Agricultural and veterinary sciences, 8-Civil engineering, 9-Industrial and information engineering, 10-Psychology.*

Table 3. Breakdown of the dataset (26.217 professors) by personal and context variables.

Variable	Level	Share
Gender	F	33.1%
	M	66.9%
Age	Less than 35	0.5%
	35-44	22.6%
	45-54	42.0%
	55-65	34.5%
	Over 65	0.4%
Univ. size	Big	65.0%
	Medium	34.1%
	Small	0.9%
Univ. location	South	28.9%
	Center	27.7%
	North	43.4%

All variables were extracted from the database of Italian professors maintained by the Minister of University and Research (MUR).[5]
For setting the bibliometric dataset, we used the author name disambiguation algorithm developed by D'Angelo, Giuffrida, and Abramo (2011), based on the coupling of the publications extracted from the Web of Science *core collection* by Clarivate Analytics and the MUR database. This algorithm assigns a WoS publication (articles, reviews, letters, and conference proceedings only) to a given professor if the latter:

- Has a name matching one of the authors in the publication byline;
- Is affiliated with one of the recognized universities listed in the publication's author addresses;
- Is associated with a discipline that aligns with the subject category (SC) of the publication;
- Was on staff as of December 31 of the year preceding the publication year.

Once we have assigned to each professor in the dataset the publications he/she has authored, we calculate two indicators, namely output (O) and fractional output (FO). The first is the simple count of the authored publications; the second is the fractional count, whereby we sum up the fractional contribution of the author to its publications, i.e., for each publication, the reciprocal of the number of co-authors

[5] For each professor this database provides information on their name and surname, gender, affiliation, discipline, field and academic rank, at close of each year. http://cercauniversita.cineca.it/php5/docenti/cerca.php, last access on 30 January 2025.

and, for publications in life science, also the co-authorship type (intramuros vs extramuros) and the position in the byline.[6]

Finally, we measure the effect of ASN in binary terms, i.e. through a dummy variable, taking the value 1 if the indicator (O or FO) measured in 2013-2017 is greater than the value measured in the previous five-year period (2008-2012); 0 otherwise.

Results

Descriptive statistics

In this section, we present the descriptive statistics derived from both the survey and bibliometric data, focusing on the perceived and measured effects of the ASN on scientific productivity. The survey data captures the perceptions of researchers regarding the effects of the ASN, while the bibliometric data reflects actual changes in productivity between two five-year periods (2008–2012 and 2013–2017). By examining the distribution of the ASN effect across gender, age groups, university size, geographical areas, and CUN disciplinary areas, we aim to highlight the alignment and discrepancies between perceived and measurable impacts of this evaluation tool in STEMM disciplines.

The proportions of researchers reporting an "ASN effect" versus "No ASN effect" from the survey differ from evidence revealed by the bibliometric analysis. The majority of respondents (65%) indicate "No ASN effect" on productivity. This suggests that most researchers perceive their productivity as not being significantly influenced by the ASN. In contrast, the bibliometric analysis shows the opposite pattern, with the "ASN effect" representing the majority (62%), reflecting measurable increases in productivity attributed to the ASN.

The following Figure 1 presents the distribution of the ASN effect ("ASN effect") and no ASN effect ("No ASN effect") by gender, based on survey and bibliometric data.

[6] For the life sciences, widespread practice in Italy is for the authors to indicate the various contributions to the published research by the order of the names in the listing of the authors. For the life science SCs publications, we give different weights to each co-author according to their position in the list of authors and the character of the co-authorship (intra-mural or extra-mural) as suggested in Abramo, D'Angelo and Rosati (2013). If the first and last authors belong to the same university, 40% of contribution is assigned to each of them, the remaining 20% is divided among all other authors. If the first two and last two authors belong to different universities, 30% of contribution is assigned to the first and last authors, 15% of the citation is attributed to the second and last authors but one, the remaining 10% is divided among all others.

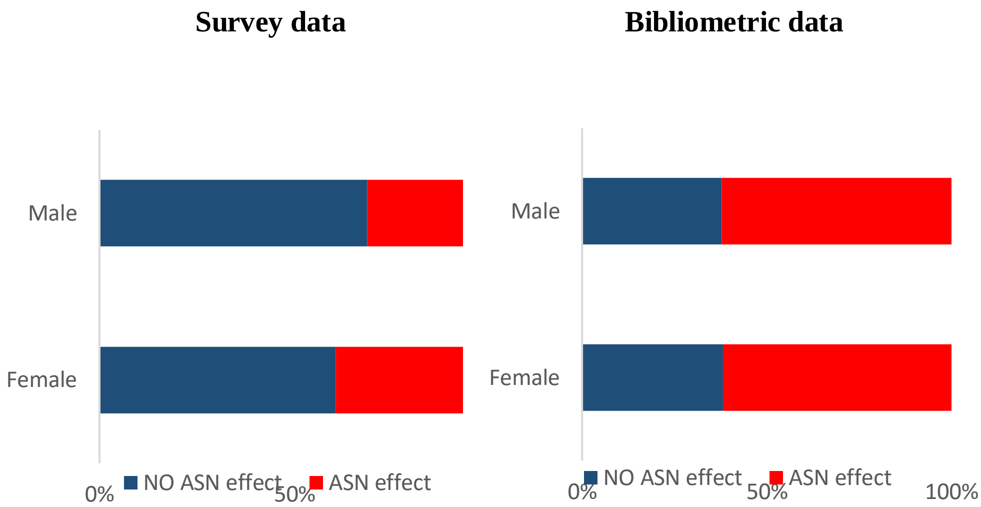

Figure 1. ASN effect by gender.

In the survey data, for both males and females, the majority report that the ASN has no effect on their scientific productivity. The bibliometric data, however, reveals a different pattern. In this case, the ASN effect appears to be more significant for both males and females, indicating a measurable increase in their scientific productivity. Figure 2 shows the distribution of the ASN effect ("ASN effect") and no ASN effect ("No ASN effect") for both survey and bibliometric data across different age groups: <35, 35-44, 45-54, 55-65, >65.

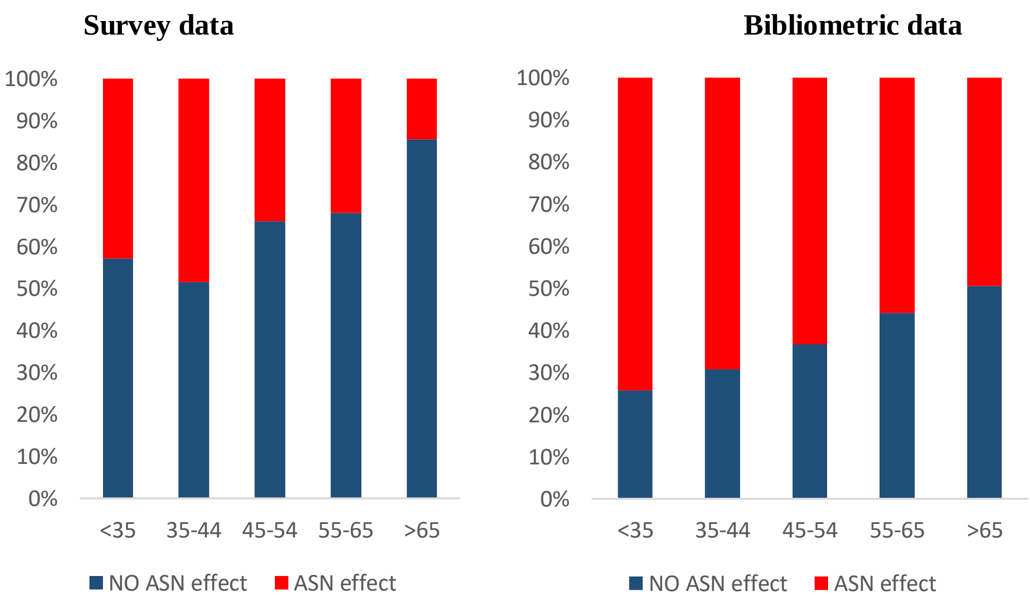

Figure 2. ASN effect by age.

In the survey-based chart, younger academics (<35 and 35–44 age groups) report a higher proportion of the "ASN effect" (red), indicating that these groups perceive a stronger impact of the ASN on their scientific productivity. The proportion of the "ASN effect" decreases progressively with age, becoming particularly small in the >65 group, where the "No ASN effect" (blue) dominates. This pattern suggests that younger researchers, who are likely at the beginning or mid-stages of their careers, feel more influenced by the ASN compared to their older counterparts. Similarly, the bibliometric-based chart (second figure) demonstrates that younger researchers (<35 and 35–44 age groups) also show the highest measurable productivity increases (red). However, a notable difference emerges in older age groups (45–54 and 55–65), where a higher proportion of the "ASN effect" is observed compared to the survey results. Even in the >65 group, a significant proportion of the "ASN effect" is evident in the bibliometric data.

Figure 3 presents the distribution of the ASN effect ("ASN effect") and no ASN effect ("No ASN effect") by geographical area (South, Center, and North).

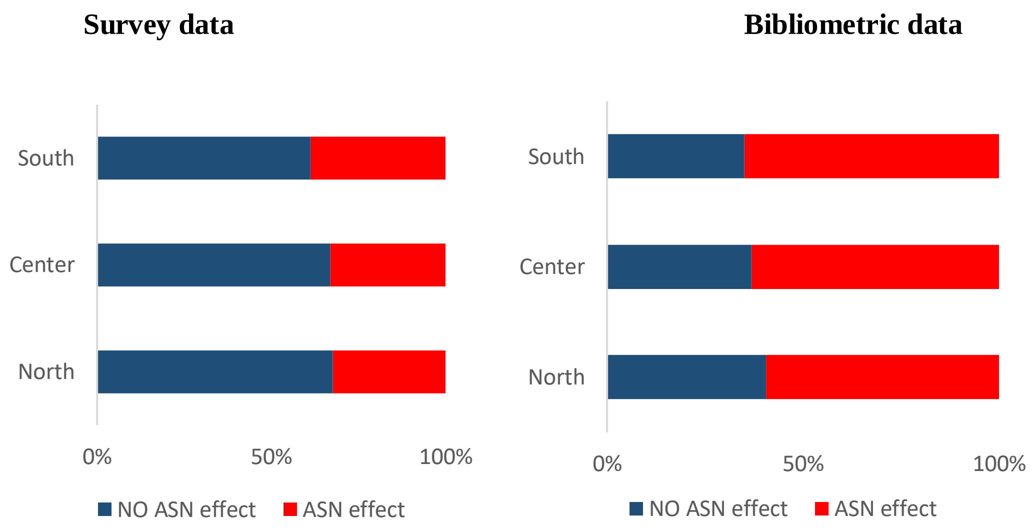

Figure 3. ASN effect by geographical area.

In the survey data, the majority of respondents indicate that the ASN has had no effect on their scientific productivity across all three regions. However, the proportion of respondents reporting an ASN effect (red) appears to be slightly higher in the South as compared to the Center and the North, suggesting that researchers in this macro-region perceive a stronger influence of the ASN on their academic output. The bibliometric data, on the other hand, present a different trend, potentially indicating a stronger measurable ASN effect across regions.

Figure 4 presents the distribution of the ASN effect and no ASN effect by university size.

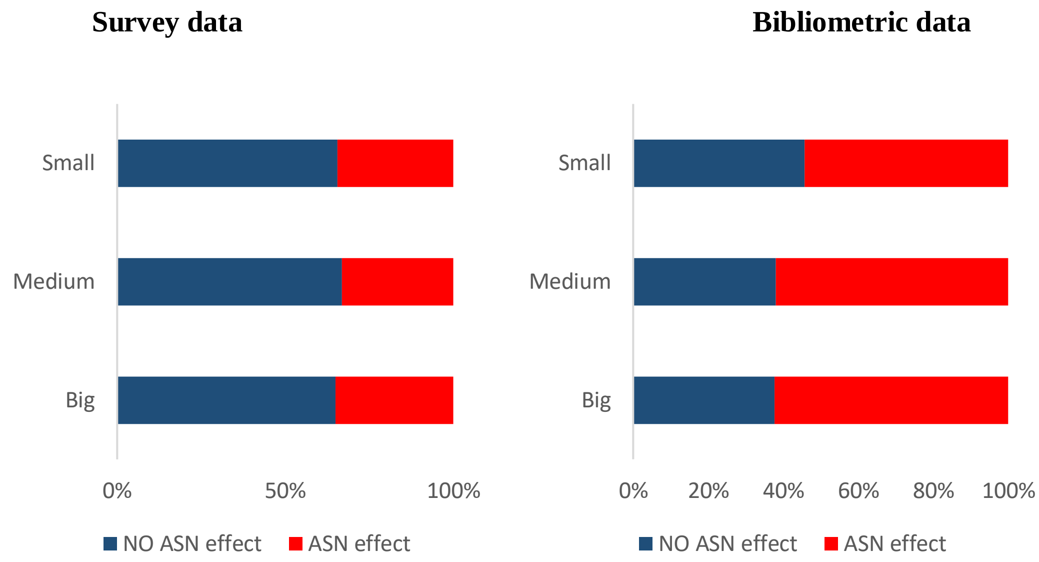

Figure 4. ASN effect by university size.

In the survey data, the majority of respondents across all university sizes (small, medium, and big) indicate that the ASN has had no effect on their scientific productivity. The bibliometric data reveals that the ASN effect appears to be more significant across all university sizes, indicating a measurable increase in scientific productivity.

Finally, Figure 5 illustrates the distribution of the ASN effect ("ASN Effect") and no ASN effect ("No ASN Effect") across different CUN categories (CUN 1 to CUN 11) for both survey-based and bibliometric data.

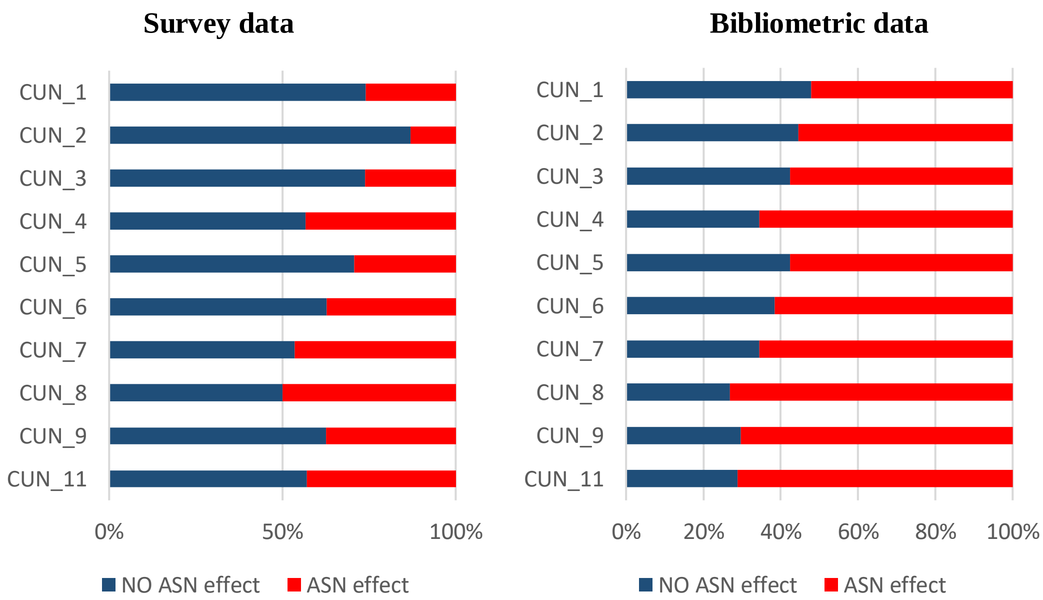

Figure 5. ASN effect by disciplinary area.

In the survey-based chart, some CUN areas, such as CUN 7 (Agricultural and Veterinary Sciences) and CUN 8 (Civil Engineering and Architecture), show a relatively high proportion of respondents reporting an "ASN Effect" (red). Conversely, fields like CUN 1 (Mathematics and Informatics), CUN 2 (Physics), and CUN 3 (Chemistry) report lower levels of perceived impact, with the "No ASN Effect" (blue) dominating. The bibliometric chart presents a different perspective on the ASN effect, reflecting changes in scientific productivity. Fields such as CUN 1 (Mathematics and Informatics) and CUN 2 (Physics) display lower proportions of the "ASN Effect" in terms of measurable increases in productivity. In contrast, CUN 8 (Civil Engineering and Architecture) shows a higher difference in productivity between the two periods (2008–2012 and 2013–2017), indicating a stronger bibliometric impact of the ASN.

The descriptive analysis highlights differences between survey-based perceptions and bibliometric evidence regarding the impact of the ASN on scientific productivity. While the majority of surveyed researchers report no significant effect of the ASN, bibliometric data suggest a measurable increase in productivity.

The econometric model

Survey

We applied a logit model, a statistical technique used to examine the relationship between a binary outcome variable and one or more predictor variables. Specifically, it models the log odds of the binary outcome as a linear function of the predictors and employs a logistic function to estimate the probability of the outcome being 1—in this case, whether the ASN influenced academics' scientific productivity. The logit model incorporates the survey's methodological design, including the sampling process. Specifically, sampling weights were included in the analysis to account for the probability of each observation being selected. These weights were also used to adjust for nonresponse and ensure that the estimates reflect the characteristics of the target population. To ensure consistency with the focus of this study, only academics who had participated in at least one ASN evaluation cycle were involved and invited to answer this question. The productivity perceived by the respondents refers to their subjective evaluation of the impact of the ASN on their scientific output, varying according to their disciplinary field. This perception includes aspects such as the volume of publications, the effort required to align with ASN standards, and the prioritization of specific research outputs.

The following table 3 presents the results of the logistic regression where "Productivity" is the dependent variable (1: ASN effect on productivity - 0: No ASN effect on productivity).

Table 3. Logistic regression on the effect of ASN (1) vs. no effect of ASN (0) on research productivity: evidence from survey data.

	Coef.	**Std Err.**	**[95% Conf. Interval]**	
Gender (1=male;0=female)	-0.330**	(0.166)	-0.655	-0.006
Age: <35	1.580**	(0.643)	0.319	2.840
Age: 35 - 44	1.820***	(0.356)	1.122	2.518
Age: 45 - 54	1.149***	(0.345)	0.473	1.826
Age: 55 - 65	1.020***	(0.341)	0.351	1.688
Univ. Medium Size vs Univ. Large Size	-0.129	(0.179)	-0.480	0.222
Univ. Small Size vs Univ. Large Size	-0.0396	(0.222)	-0.475	0.395
Geo: North vs Center	-0.100	(0.195)	-0.482	0.282
Geo: South and Islands vs Center	0.235	(0.215)	-0.187	0.657
Cun Area 1 vs Cun Area 2	0.834*	(0.478)	-0.102	1.770
Cun Area 3 vs Cun Area 2	0.801	(0.495)	-0.170	1.771
Cun Area 4 vs Cun Area 2	1.751***	(0.557)	0.658	2.843
Cun Area 5 vs Cun Area 2	1.061**	(0.440)	0.198	1.923
Cun Area 6 vs Cun Area 2	1.547***	(0.435)	0.695	2.399
Cun Area 7 vs Cun Area 2	1.798***	(0.454)	0.907	2.688
Cun Area 8 vs Cun Area 2	1.912***	(0.513)	0.906	2.918
Cun Area 9 vs Cun Area 2	1.405***	(0.434)	0.555	2.255
Cun Area 11 vs Cun Area 2	1.680***	(0.577)	0.550	2.811
Constant	-2.871***	(0.553)	-3.954	-1.788

*** p<0.01, ** p<0.05, * p<0.1

The regression results highlight several important patterns regarding the factors influencing perceptions of productivity increases attributed to the ASN. Gender plays a significant role, with male academics being less likely than their female counterparts to report that the ASN has positively impacted their productivity. Age also emerges as a crucial factor, with younger academics, particularly those under 35 and in the 35–44 age range, significantly more likely to report productivity increases due to the ASN. This suggests that early-career researchers, who are often more dependent on evaluation systems for career progression, are more responsive to the pressures and incentives created by the ASN. However, senior academics, while slightly less likely to attribute an impact compared to younger colleagues, also report the effects of the ASN on their productivity. This indicates that the influence of the ASN is not limited to any single career stage but is felt across all age groups, underscoring its pervasive impact on academic publishing behaviors.

The results show considerable variation when considering disciplinary differences (as represented by CUN areas). Academics in certain fields, such as those in Areas 4 (Earth Sciences), 5 (Biology), 6 (Medicine), 7 (Agricultural and Veterinary Sciences), 8 (Civil Engineering and Architecture), and 9 (Industrial and Information Engineering), are significantly more likely to attribute productivity increases to the ASN compared to those in Area 2 (Physics). This suggests that disciplines with

different publication practices and evaluation standards may respond differently to the incentives of the ASN, with some fields feeling a stronger push to align their outputs with its requirements.
Geographic location and university size do not show a significant effect on the likelihood of reporting productivity increases attributed to the ASN. This suggests that the likelihood of reporting productivity increases attributed to the ASN appears to be primarily influenced by individual characteristics or specific area-based indicators rather than by institutional factors.

Bibliometric analysis

The logistic regression model presented here examines the factors influencing the likelihood of observing a measurable increase in scientific productivity attributed to the ASN, as determined by bibliometric data. The dependent variable is binary, taking the value of 1 if there is a measurable increase in productivity between the two periods (2008–2012 and 2013–2017) and 0 otherwise.
Table 4 presents the results of the logistic regression, characterized by the following features.
Number of obs = 26217
Wald chi2(18) = 742.25
Prob > chi2 = 0.0000
Log pseudolikelihood = -17013.07
Pseudo R2 = 0.0222

Table 4. Logistic regression on the effect of ASN (1) vs no effect of ASN (0) on research productivity: evidence from bibliomrtric data.

	Coef.	**Std Err.**	**[95% Conf. Interval]**	
Gender (1=male;0=female)	0.015	0.029	-0.041	0.071
Age: <35	1.100***	0.288	0.535	1.665
Age: 35 - 44	0.841***	0.21	0.428	1.253
Age: 45 - 54	0.578***	0.209	0.167	0.988
Age: 55 - 65	0.253	0.209	-0.157	0.663
Univ. Medium Size vs Univ. Large Size	-0.062**	0.028	-0.116	-0.007
Univ. Small Size vs Univ. Large Size	-0.417***	0.14	-0.692	-0.142
Geo: North vs Center	-0.214***	0.032	-0.275	-0.152

Geo: South vs Center	0.026	0.035	-0.042	0.095
Cun Area 1 vs Cun Area 2	-0.871***	0.087	-1.042	-0.7
Cun Area 3 vs Cun Area 2	-0.679***	0.093	-0.86	-0.497
Cun Area 4 vs Cun Area 2	-0.647***	0.088	-0.82	-0.475
Cun Area 5 vs Cun Area 2	-0.269**	0.108	-0.48	-0.057
Cun Area 6 vs Cun Area 2	-0.591***	0.084	-0.756	-0.427
Cun Area 7 vs Cun Area 2	-0.370***	0.081	-0.529	-0.211
Cun Area 8 vs Cun Area 2	-0.297***	0.088	-0.47	-0.123
Cun Area 9 vs Cun Area 2	0.055	0.102	-0.144	0.254
Cun Area 11 vs Cun Area 2	-0.104	0.085	-0.271	0.063
Constant	0.490**	0.222	0.054	0.926

*** p<0.01, ** p<0.05, * p<0.1

Observing the results of bibliometric regression, we see that Age emerges as a strong predictor, with younger academics, particularly those under 35 and in the 35–44 age range, significantly more likely to report the outcome under consideration compared to older colleagues. While the likelihood decreases with age, academics aged 45–54 also show significant effects. However, for those aged 55–65, the effect is no longer statistically significant, suggesting that the influence of this factor diminishes with seniority.

Institutional size plays an important role, with academics affiliated with medium-sized and small universities being less likely to report the outcome compared to those at large universities. This effect is particularly pronounced for small universities, where the likelihood of reporting the outcome is significantly reduced. These findings indicate that institutional environments at larger universities may create easier conditions for achieving the specified outcome.

Geographic differences also emerge, with academics in the North of Italy being significantly less likely to report the outcome compared to those in the Center. However, no significant differences are observed between the South and Center, suggesting a more uniform experience in those regions.

The results reveal considerable variation across disciplinary areas. Academics in Areas 1 (Mathematics and Informatics), 3 (Chemistry), 4 (Earth Sciences), 5 (Biology), 6 (Medicine), 7 (Agricultural and Veterinary Sciences), and 8 (Civil

Engineering and Architecture) are significantly less likely to report the outcome compared to those in Area 2 (Physics). Notably, Area 9 (Industrial and Information Engineering) and Area 11 (only Psychology sub-area) do not show significant differences compared to Area 2, suggesting closer alignment in these disciplines.
The comparison between the survey and bibliometric analyses reveals both convergences and divergences in the factors shaping the impact of the ASN on scientific productivity in STEMM fields. Both approaches underscore the strong influence of age, with younger academics, particularly those under 35 and in the 35–44 age range, significantly more likely to report productivity increases attributed to the ASN. This suggests that early-career researchers are more responsive to the ASN's incentives for career progression. Institutional size, however, emerges as a significant factor only in the bibliometric analysis, where academics at medium-sized and small universities report fewer productivity increases compared to their counterparts at larger universities. This likely reflects disparities in resources, access to academic knowledge networks and research infrastructure. These constraints can make it harder to align with ASN-driven incentives, particularly in fields where collaboration and resource intensity are critical for publishing high-quality work.
Regarding the geographic location, the bibliometric analysis identifies lower effects in the North of Italy, while the survey finds no significant regional differences. Both analyses highlight disciplinary differences, although in contrasting directions: the survey identifies stronger effects in fields such as Earth Sciences, Biology, and Medicine, while physicists report being less influenced by the ASN, suggesting that their perceived increase in productivity is less tied to the evaluation tool. In contrast, the bibliometric analysis indicates that Physics, taken as the baseline category in the model, shows higher productivity increases compared to other disciplines. This discrepancy suggests that while physicists do not attribute their increased productivity to the ASN in the survey, the bibliometric evidence points to an actual increase in their output, which may instead be driven by other factors, such as intrinsic disciplinary dynamics and stronger collaboration networks.

Conclusions

This study highlights a significant misalignment between researchers' perceptions of productivity increases attributed to the Italian Scientific Habilitation (ASN) and the evidence obtained through bibliometric assessments. While the bibliometric analysis reveals that the majority of academics (62%) experienced measurable increases in scientific productivity following the introduction of the ASN, survey data indicate that most researchers (65%) perceive little to no effect on their productivity. This discrepancy underscores a fundamental difference in how the effects of the evaluation systems are experienced versus their quantifiable outcomes.
The divergence between perception and evidence is particularly notable across demographic and contextual factors. Younger researchers and those at earlier stages of their careers are more likely to report productivity increases, both in survey responses and bibliometric data, reflecting their stronger dependence on evaluation systems for career progression. However, in older age groups, while bibliometric evidence points to measurable productivity increases, these are often not recognized

or attributed to the ASN by the researchers themselves. Similarly, disciplinary differences reveal contrasting patterns: researchers in fields such as Medicine and Engineering report and exhibit higher productivity increases, while those in Physics and Mathematics show a significant bibliometric impact but perceive less influence from the ASN.

These differences have critical implications for researchers, institutions, and policymakers. As it regards researchers, the key question is: what lies at the root of the mismatch between perception and evidence? If the discrepancy stems from researchers failing the habilitation exercises, it could significantly undermine trust in evaluation systems, particularly among those who feel their contributions are undervalued or overlooked. In such cases, researchers might abandon efforts toward continuous improvement or resort to counterproductive behaviors. These could include prioritizing short-term outputs over long-term discoveries or interdisciplinary work, or engaging in unethical practices like excessive self-citation, citation networks, salami-slicing publications, or searching for honorary authorship. As for institutions, universities must navigate the varying impacts of evaluation systems across disciplines and demographics. The observed disparities may suggest that a one-size-fits-all approach to research assessment is insufficient. Institutions should aim to foster environments where diverse academic contributions, including teaching, mentorship, and technology transfer, are valued alongside publications.

Talking about policymakers, the findings emphasize the need for more nuanced and inclusive evaluation policies. Efforts to improve the transparency and communication of evaluation criteria and results could help bridge the gap between perception and evidence, enhancing the legitimacy of these systems and forging researchers' virtuous behavior.

Policy recommendations stemming from this study include but are not limited to i) tailoring discipline-specific metrics that align with the unique publication practices and priorities of each field; ii) promoting transparency by clearly communicating how metrics are used in the evaluation and providing feedback to researchers on how their work aligns with institutional and national goals; and iii) addressing the equity gaps by implementing targeted measures to reduce disparities observed in gender, geographic location, and institutional size, ensuring fair and equitable evaluation processes.

Finally, this research underscores the need for ongoing dialogue among policymakers, institutional leaders, and researchers to ensure that evaluation systems align with academic values and societal goals. By bridging the gap between perception and evidence, we can foster trust, inclusivity, and innovation in the academic community, ensuring that research evaluation serves its ultimate purpose: advancing knowledge and addressing global challenges.

Despite its contributions, this study has methodological limitations. The survey data relies on self-reported perceptions, which may be influenced by personal biases or an incomplete understanding of the factors driving productivity changes. Conversely, bibliometric analyses rely on proxies for productivity, which may overlook qualitative aspects of academic work. Furthermore, bibliometric analyses infer causality based on observed trends, which may not fully capture the complex interplay of motivations and constraints affecting researchers. The limitations of comparing the possible mismatch between perceptions and bibliometric evidence are twofold. On the one hand, the strength of the causal attribution of changes in research productivity to the ASN is not the same. It derives from individual appreciation in the case of the survey, while it is inferred in the case of the bibliometric analysis by observing the levels and characteristics of productivity in the different fields before and after the introduction of the ASN. On the other hand, the survey also collects the perceptions of the respondents on the importance of other factors beyond the ASN on the changes in research productivity. Therefore, the attribution of the effect observed to the ASN can be calibrated with respect to other causes that played a role in the production of the effect.

Future research should explore the underlying reasons for these misalignments, incorporating mixed methods and longitudinal designs to better understand the evolving relationship between perception, evidence, and the broader academic environment.

In conclusion, addressing the gap between researchers' perceptions and bibliometric evidence is essential for building trust and ensuring that evaluation systems serve their intended purpose. By aligning these systems with academic values and promoting inclusivity, we can foster environments that support both individual and collective advancement in knowledge creation.

References

Abramo, G. (2024). The forced battle between peer-review and scientometric research assessment: Why the CoARA initiative is unsound. *Research Evaluation*, rvae021, doi.org/10.1093/reseval/rvae021.

Abramo, G., & D'Angelo, C.A. (2021). The different responses of Italian universities to introduction of performance-based research funding. *Research Evaluation,* 30(4), 514–528.

Abramo, G., & D'Angelo, C.A. (2014). How do you define and measure research productivity? *Scientometrics*, 101(2), 1129-1144.

Abramo, G., D'Angelo, C.A., & Rosati, F. (2013). Measuring institutional research productivity for the life sciences: the importance of accounting for the order of authors in the byline. *Scientometrics,* 97(3), 779-795.

Archambault, É., Vignola-Gagné, É., Côté, G. et al. (2006). Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics* 68, 329–342.

Bornmann, L., & Daniel, H. D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64(1), 45–80.

Brembs, B., Button, K., & Munafò, M. (2013). Deep impact: Unintended consequences of journal rank. *Frontiers in Human Neuroscience*, 7, 291.

de Diego, I.M., Prieto, J.C., Fernández-Isabel, A., Gomez, J. & Alfaro, C. (2024). Framework for scoring the scientific reputation of researchers. *Knowledge and Information Systems*, 66, 3523–3545.

de Rijcke, S., Wouters, P.F., Rushforth, A.D., Franssen, T.P., & Hammarfelt, B. (2016). Evaluation practices and effects of indicator use—a literature review. *Research Evaluation*, 25(2), 161–169.

DORA. (2012). Declaration on Research Assessment.

Fire, M., & Guestrin, C. (2019). Over-optimization of academic publishing metrics: Observing Goodhart's Law in action. *GigaScience*, 8(6), giz053.
Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science*, 178(4060), 471–479.
Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature*, 520(7548), 429–431.
Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46), 16569–16572.
Ioannidis J.P.A., & Maniadis, Z. (2023). In defense of quantitative metrics in researcher assessments. *PLoS Biology*, 21(12): e3002408.
Larivière, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. R. (2013). Bibliometrics: Global gender disparities in science. *Nature News*, 504(7479), 211.
McKiernan, E. C., et al. (2016). How open science helps researchers succeed. *eLife*, 5, e16800.
Moed, H. F. (2005). *Citation analysis in research evaluation*. Springer Science & Business Media.
Moher, D., Naudet, F., Cristea, I. A., Miedema, F., Ioannidis, J. P., & Goodman, S. N. (2022). Assessing scientists for hiring, promotion, and tenure. *PLOS Biology*, 20(3), e3001606.
Mongeon, P., Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, 106, 213–228.
Muller, J. Z. (2018). *The tyranny of metrics*. Princeton University Press.
Nicholas, D., Watkinson, A., Boukacem-Zeghmouri, C., Rodriguez-Bravo, B., Xu, J., Abrizah, A., & Herman, E. (2017). Early career researchers and their publishing and authorship practices. *Learned Publishing*, 30(2), 101–112.
Sahel, J. A. (2011). Quality versus quantity: Assessing individual research performance. *Science Translational Medicine*, 3(84), 84cm13.
Seeber, M., Cattaneo, M., Meoli, M., & Malighetti, P. (2019). Self-citations as strategic response to the use of metrics for career decisions. *Research Policy*, 48(2), 478-491.
Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, 3(9), 160384.
Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, 10(2), 365–391.