ASSESSING SCIENTIFIC RESEARCH & INNOVATION

An Analysis of Research Productivity Metrics and Usage Guidelines

Centre for CIVIL SOCIETY
SOCIAL CHANGE THROUGH PUBLIC POLICY

FAST INDIA

manipalfoundation

Assessing Scientific Research & Innovation:

An Analysis of Research Productivity Metrics and Usage Guidelines

First published in 2022
by Centre for Civil Society

Cover design by
Ravi Kumar Yadav,
Centre for Civil Society

Authored by
Pratyaksha Jha, Centre for Civil Society

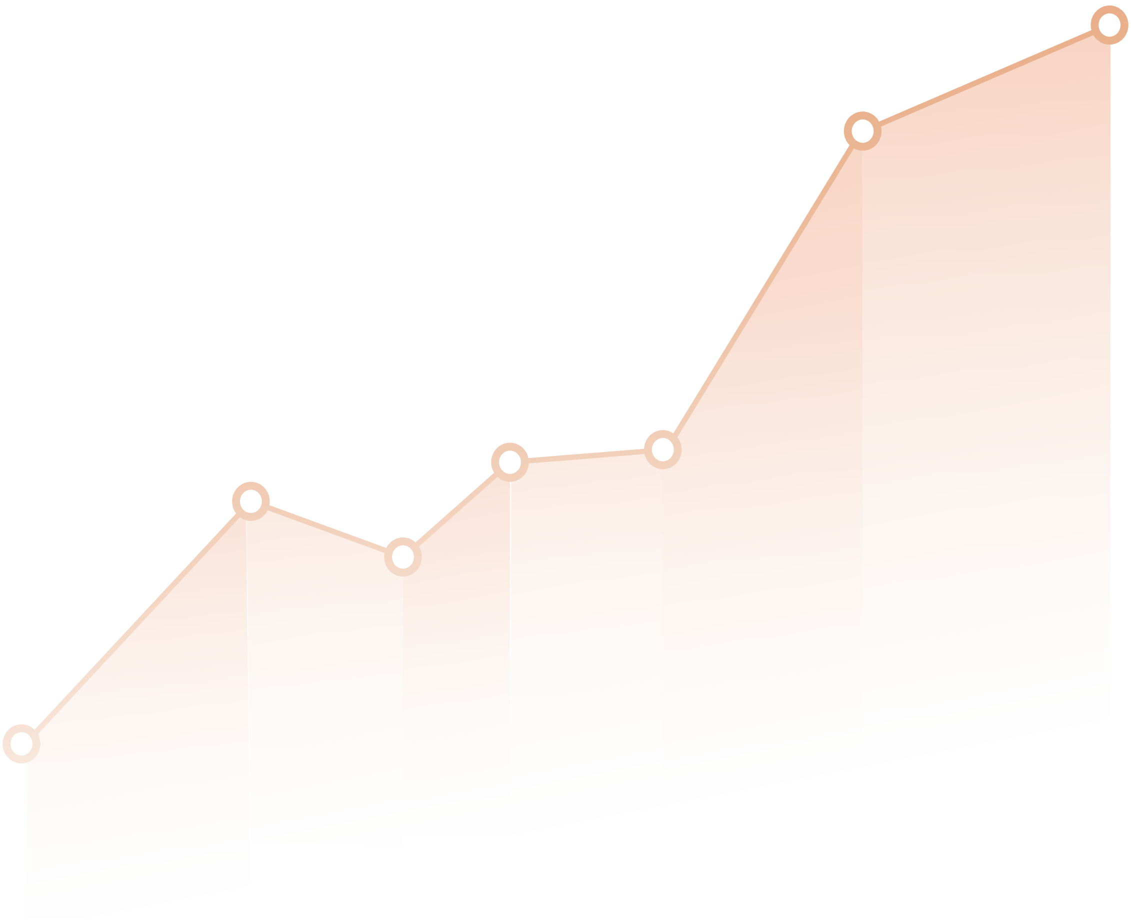
Please cite the work as follows:
Jha, Pratyaksha. 2022. *Research Productivity Measurement: An Analysis of Concepts, Methods, and Usage Guidelines.* Centre for Civil Society

Supported by
FAST India and Manipal Foundation

For more information and other requests, write to:
Centre for Civil Society
A-69, Hauz Khas, New Delhi – 110016
Phone: +91 11 26537456
Email: ccs@ccs.in | Website: www.ccs.in

ASSESSING SCIENTIFIC RESEARCH & INNOVATION

An Analysis of Research Productivity Metrics and Usage Guidelines



Centre for CIVIL SOCIETY
SOCIAL CHANGE THROUGH PUBLIC POLICY

FAST INDIA

manipalfoundation

CONTENTS

Abbreviations

API	Application Programming Interface
DORA	Declaration on Research Assessment
ENID	European Network of Indicator Developers
FSS	Fractional Scientific Strength
ICTs	Information and Communication Technologies
JIF	Journal Impact Factor
OECD	Organisation for Economic Cooperation and Development
pre-SCORE	Peer Review Evaluation Score
R&D	Research and Development
ROAI	Relative Open Access Index
SCI	Science Citation Index
STI	Science, Technology, and Innovation
TARA	Tools to Advance Research Assessment

Executive Summary

Research productivity refers to the efficacy of research processes with respect to the quantity and quality of its contributions. In the process of building our three-part series on 'assessing scientific research and innovation', we learned that research productivity assessments are a tool used for many processes in the STI (Science, Technology, and Innovation) ecosystem. These processes include research funding allocation, and evaluating the efficacy of scientific R&I (Research & Innovation) in academic institutions. This report reviews contemporary approaches to assessments of scientific research productivity, and highlights where interventions can be made to make them more holistic and accurate.

Scientific research productivity is a concern that affects diverse stakeholders such as government agencies, industrial bodies, and academic groups. As a result, existing scholarship in this area is vast, but fairly disparate. We aim to build a resource that brings together concepts used in its assessment and examines their application. In the first chapter, we outline the usage and limitations of four prominent research productivity assessment tools— bibliometrics, open access metrics, peer review metrics, and economic productivity oriented metrics. The second chapter examines international frameworks that set out guidelines for assessment exercises, and concludes with a comment on their efficacy. Our final chapter analyses two key areas where future interventions can be made: (a) problems of standardisation and yielding comparative research productivity data, and (b) the need to make research productivity metrics more relevant to the economic and social contexts of developing countries.

We conclude by offering insights on how gaps in these two areas can be addressed. Our focus here is on how greater standardisation can be achieved without compromising on the imminent need to build more context-sensitive assessment tools. Our inferences include:

- The creation of more international guidelines on standardisation would help produce more comparative research productivity data.

  Establishing normative standards around how to use different metrics would help make their usage more accountable as well as accessible.
- There is also an imminent need to create frameworks specific to developing countries, detailing research productivity measures that further shared goals based on economic development.
- Measures such as journal impact rankings and citation impact should be made more accommodating of researchers who lack resource privilege.
- Resource constraints that prevent developing countries from accessing particular forms of research productivity success need to be acknowledged and addressed.

Introduction

Measurements of productivity are focused on understanding the relationship between inputs and outputs linked to particular processes. They typically seek to examine the ability of input-process-output chains to meet certain end-goals or parameters that denote success, efficiency, or quality. Insights generated from measuring productivity are used to define future imperatives and carry out strategic interventions to improve processes. Insights thus generated from measuring productivity are then deployed to define future imperatives and interventions within the scope of what is being measured, and to improve systems and processes. Research productivity measurement focuses on applying these parameters to research processes, outcomes, and outputs, typically pertaining to particular institutions, regions, or disciplines.

Broadly, research productivity can be disambiguated in terms of aspects such as quantity of research output, quality of research contributions, impact of research insights, or the incorporation of particular ethical or normative standards into research processes. At the outset, this indicates that what is denoted by research productivity can itself be defined in multiple, and sometimes overlapping ways. Given that research productivity can be denoted in terms of so many different parameters, it is also difficult to find aggregated data points that are able to take all of these into account. As a result, research productivity measures can employ quantitative as well as qualitative tools, and larger measurement exercises, such as evaluating an entire country's R&D output typically apply more than one method to cover a broader range of parameters.

It is also important to note that many situations call for the evaluation of research outcomes and processes without explicitly categorising the processes involved as research productivity measurement. Any data that is used to make inferences about the value of research being generated, constitutes a measure of research productivity. Then, each such measurement exercise inevitably pertains to its own specific parameters of what denotes utility and value. While in the case of R&D for industrial output, economic growth is likely to be correlated with success, the process of allocating research funding would plausibly function on a calculus of maximising both quality and quantity of research simultaneously. Therefore, the research productivity data yielded from measurement exercises exists in many different forms.

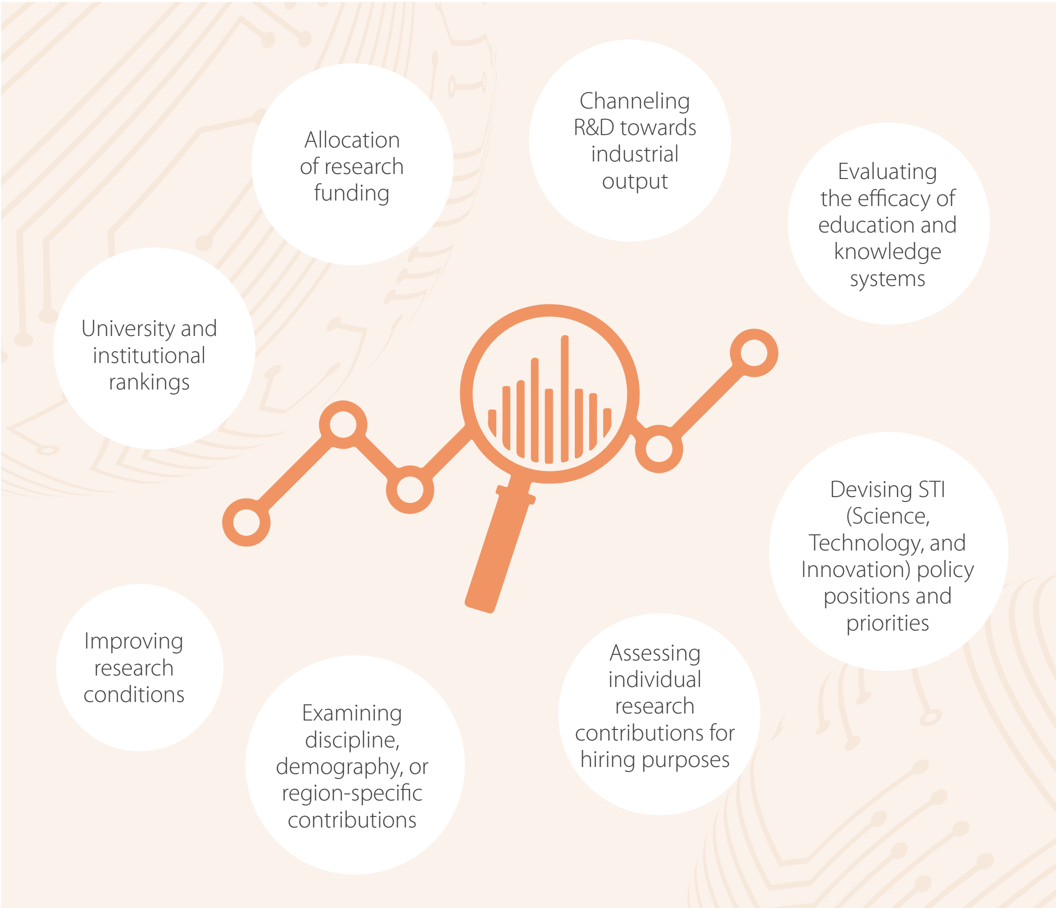


Figure 1: Prominent uses of Research Productivity Assessments

This report aims to address literature in the field of scientific research productivity by reviewing the inadequacies in how it is assessed in many parts of the world. To this end, it reviews commonly used research productivity metrics, as well as international frameworks that provide guidelines on assessment exercises. Consequently, it identifies two thematic areas where gaps and limitations lie— namely, the standardisation of assessment data, and underrepresentation of the needs of developing countries in existing tools.

It is the second report in our three-part study titled *Assessing Scientific Research and Innovation*. The first report in our series, *Study of frameworks and parameters for evaluating institutional research*, examines different indices used to measure scientific innovation in Indian institutions of higher education (CCS 2022). The third report in this series, titled *The South Asian Case*, will focus on specific interventions that can be used to make research assessments more relevant and accessible for use in South Asian countries.

Section 01

Types of Metrics

This section selects four prominent types of metrics that recur in research productivity measurement literature— bibliometrics, open access indicators, peer review, and economic productivity-oriented measures— and describes their usage as well as limitations. It also highlights more nascent indicators by way of altmetrics and measures of research equity. The purpose behind this inquiry is to examine the efficacy of these tools in measuring different aspects of research productivity.

1.1 Bibliometrics

A. Definitions

Bibliometric measures are quantitative research productivity metrics that focus on research outputs in the form of journal articles, research papers, and other forms of written research output. Bibliometrics are defined as "statistical or mathematical method(s) for counting the number of academic publications, citation and authorship" (Directorate-General for Research 2010). They encompass the following types of metrics (University of Waterloo Working Group on Bibliometrics 2016):

- **Publication Counts —** *Absolute number of publications*
- **Citation Impact —** *Absolute number of times that a given research output is cited (citation databases like Elsevier Scopus, Thomson Reuter's Web of Science, and Google Scholar are typically used to measure citation impact)*
- **H-index —** *A researcher's h-index is denoted by the largest number x such that they have at least x number of research publications that have x number of citations (variations based on generalisations of the h-index, using different parameters include the g-index, the m-index, the i10-index, and the Py-index)*
- **Collaboration Networks —***Measures of types and degrees of researcher collaboration, through parameters like co-authorship, industrial collaborations, and international collaborations*
- **Journal Impact Ranking —** *Aggregation of citation data for journals that measures the relative importance of a particular journal (Thomson Reuter's JIF— Journal Impact Factor— is a popular example of this kind of measure)*
- **Top Percentiles —** *Measures of the most cited research outputs in particular subject areas, document types, years etc.*

Through a quantification of research output, bibliometric assessments tend to focus on drawing out insights along the lines of— (a) which journals are prominent in terms of carrying research on particular subject areas, (b) what research trends in particular disciplines look like based on sub-disciplinary thematic areas and keywords, or (c) which countries, regions, and authors are leading research output in particular fields. A large number of studies employing these measures employ data acquired from the citation databases mentioned earlier— Elsevier Scopus, Web of Science, and Google Scholar.

B. Limitations

Several limitations of bibliometric measures become important to take into account while using them:

- **The utility of the data they arrive at is heavily dependent on the database(s) used to yield it.** Therefore, it is susceptible to disciplinary, linguistic, and other gaps that might exist within the database. This limitation is compounded by the fact that there is little scope to holistically evaluate whether or not a database is itself bias-free and fully representative of research in a particular field.
- **Citation impact can realistically be measured only after a time lapse following the publishing of research.** If one were to measure the citation impacts of two research papers such that one was released several decades after the other, the amount of time since publication could play a central role in determining the number of times either paper has been cited.
- **Citation impact does not account for human biases in how research output is received and cited.** Structural inequalities in terms of class, gender, race, and ethnicity are not taken into account by this measure. These could hold a number of implications for researchers, ranging from a lack of resources to access high quality research infrastructure or cover costs of research and publication, to other researchers displaying behavioural biases against citing their work. Presumably then, researchers having disprivileged social locations are likely to fare worse on this metric. Without room to account for why patterns in citation could favour researchers from particular demographics, citation-related metrics provide inaccurate insights into individual research productivity.
- **Citation impact does not make distinctions based on the nature of the citations being counted.** A publication that is cited multiple times to be debunked cannot be differentiated from one that is cited as a credible, influential source.
- **Parameters linked to publication counts can, at best, make a claim about the quantity of research being produced.** They are unable to validate its quality, and prima facie do not offer any strong verifiable claims about how research quantity and quality are correlated.
- **Parameters such as author collaboration and journal impact ranking are based on certain assumptions around value and quality.** While author collaboration is context-driven to the extent that it could have different implications within different fields, journal impact rankings generalise the individual quality of research articles based on how they assess the credibility of journals. Either indicator is susceptible to biases linked to unequal access to opportunities on researchers' part.

Based on these considerations, it emerges that most limitations of bibliometric measures converge on there being insufficient room to account for context-sensitive qualitative parameters. While these limitations do not invalidate the role of bibliometrics or make it entirely redundant, they do point towards a need to use other methods alongside them to get a holistic and accurate picture of research productivity.

ALTMETRICS

Altmetrics , or '**alternative metrics**', focus on measuring the impact of research through an engagement with its online footprint. These newer forms of measurement focus on the transition from print-mode research that had journals, peer reviews, and formal citations at its heart, towards online platforms like Twitter and scholarly blogs (Priem et al 2010). Using public APIs (application programming interfaces) to gather data, they measure engagement based on semantic content such as usernames, timestamps, and tags (ibid.). While initiatives to build more detailed, consolidated frameworks and guidelines on altmetrics and their usage are underway, there could be a lack of quality control in their absence. Without proper parameters that define what research productivity looks like in altmetric data terms, one runs the risk of conflating unrelated forms of digital engagement with knowledge dissemination.

Figure 2: Altmetrics

1.2 Open Access

A. Definitions

Open access literature, one of the most widely discussed aspects of a broader movement around open science, refers to literature that is digital, online, free of charge, and free of most copyright and licensing restrictions (Peter Suber 2004). The Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003), signed by more than 700 research organisations to date, specifies two conditions that must be satisfied by a publication for it to be considered open access—

- that its authors and holders grant all users a free, irrevocable, worldwide right of access to it, as well as a licence to copy, use, distribute, transmit, and display it
- that a complete version of the work, as well as supplemental materials including the permissions specified by the first condition, be included in at least one online repository maintained by an academic institution, scholarly society, government agency or other credible organisation

While open access is formally defined in these terms, it is also used somewhat fluidly to denote other types of published academic material that are at times free to access and distribute — for instance, articles in repositories that are free to read but not to reuse, and articles that are published in paid journals but are open-licensed after authors pay an article processing charge (Piwowar et al. 2018).

Within the research productivity assessment paradigm, there is an increasing focus on indicators based on open access. International frameworks reviewed later on in this study posit that open sharing of scientific research data is crucial to widening the impact of research. An understanding that open access is an important research productivity parameter, especially in the age of rapid digitalisation, is reflected in the development of indices and metrics that are used to evaluate it. The Relative Open Access Index (ROAI) is one such country-level index, which measures a country's share of open access articles against the global share, assigning scores between 0 and 1 (Elango, Oh and Rajendran 2021). Similarly, there is growing scrutiny on bibliometric assessment tools due to commonly used citation databases not having enough open access repositories and publications within their scope.

OPEN SCIENCE

The term "**open science**" was coined in 2003, and denotes the understanding shared by some schools of economics that scientific knowledge generated through publicly-funded research is a public good. (OECD 2015). Open access to research output forms a significant chunk of open science. However, its scope extends beyond them to include aspects of the research process rather than being limited to its outcomes.

Open science is envisaged in terms of all components of research, at every stage of the research process, being accessible to other research practitioners (Marcus Hanwell, n.d.). It encompasses open access, open data, open source for softwares, open standards of knowledge dissemination, and exploratory features such as open peer review, where reviewers attach their names to their reviews, and open notebook science, where researchers make their notes and notebooks during the research process publicly accessible. At its heart, open science recognises the growing role of ICTs (Information and Communication Technologies) in enabling a scale of knowledge dissemination that has not been possible before. Its goals can be summarised as follows (OECD 2015):

- optimising research processes by allowing multiple researchers to work on the same data
- allowing more opportunities for participation in research, especially from those who face structural barriers that prevent them from getting access to data and resources that enable high quality research output
- putting scientific research to greater scrutiny by making it more accessible
- providing firms and individuals with greater access to scientific research, and thus incentivising the creation of more products and services based on its insights

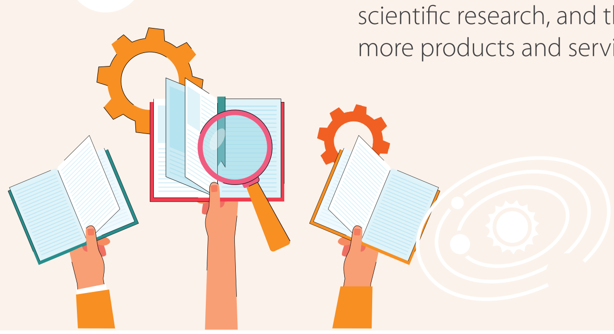


Figure 3: Open Science

B. Limitations

Even though open access-related metrics are gaining prominence within research productivity discourses, they are subject to several limitations:

- **The novelty of open access practices poses constraints for ensuring quality control while selecting open access publications and repositories.** While several attempts have been made to create standards around what constitutes a quality open access resource (as mentioned above), there exist ambiguities around when these resources can be considered credible, academically robust, and indeed, even open. This not only means that there is a lack of emphasis on measuring the quality of these resources, but also that broad contested definitions of what constitutes open access are likely to yield disparate, non-standard data on its research impact.
- **The inability to monetise open access resources makes it harder to align research with profit-oriented economic incentives, and poses contradictions with economic growth through research output.** With free, not-for-profit dissemination of knowledge being central to their nature, open access publications are typically unable to offer even small-scale monetary incentives to researchers, often relying on them to cover costs of publication instead. Further, when it comes to research that may hold economic utility for private or even public sector companies, free distribution can act to the detriment of it being put to use by them, due to the research no longer posing novelty or a competitive advantage over other companies in the same sector. While conversation on open access often emphasises publicly funded research, the lines between public and private research are far more difficult to draw in the context of developing countries, where the degrees to which private sector organisations are able to function independent of public sector investment and involvement vary significantly more than their developed counterparts.
- **There has been significant scrutiny on the growing number of predatory journals, ie. journals that charge money from researchers for publishing their research out of skewed profit incentives, and typically do not hold themselves to high standards of academic credibility while accepting submissions.** The relationship predatory journals have with open access remains contested. Some scholars posit that open access has been incorrectly accused of causing the rise of predatory journals, analysing the popular Beall's List of Potential Predatory Journals and Publications from 2012 to show how it over-generalises the role of open access in their creation (Krawczyk and Kulczyki 2021). However, it would be reasonable to assume that the lack of robust standardisation around open access resources discussed earlier makes it easier for predatory journals to wield them to produce research whose quality is under-verified and publications that operate on perverse incentives.

The key limitations that come up while using open access metrics to measure research productivity converge on a lack of standardised definitions and processes that map the quality of open access publications. As such, measures taken towards creating such standardisation would help manoeuvre through them more effectively.

1.3 Peer Review

A. Definitions

Peer review links expert qualitative perspectives on research to assessments of its quality. Its relevance is typically justified in terms of its ability to accord validity and credibility to discipline-specific research processes and insights. This measure is most often used to assess the quality of individual research contributions when they are being considered for publication in academic journals. When used thus, the peer review method has two key purposes— (1) to ensure that only high quality research is published, based on its validity, significance, and originality and (2) to work on the research draft in order to improve its quality (Kelly et al. 2014).

Peer review links expert qualitative perspectives on research to assessments of its quality. Its relevance is typically justified in terms of its ability to accord validity and credibility to discipline-specific research processes and insights. This measure is most often used to assess the quality of individual research contributions when they are being considered for publication in academic journals. When used thus, the peer review method has two key purposes— (1) to ensure that only high quality research is published, based on its validity, significance, and originality and (2) to work on the research draft in order to improve its quality (Kelly et al. 2014).

A qualitative study that surveyed 72 peer reviewers to understand the peer review method finds several points of consensus on what the process should look like (Tercier and Callabam 2007). In terms of practices, 92% reviewers reported going through a research manuscript two to three times in different sittings while peer reviewing it, asking questions about its contents in accordance with the instructions given to them by journal editors (ibid.). The core principles of peer review were identified to be (ibid.):

- effectiveness- promoting progress by improving the quality of research (74%)
- sensitivity and specificity- ensuring that relevant research is published, and poor research is not (59%)
- fairness- being objective and unbiased in one's review (28%)
- efficiency- process being timely for the author (22%) as well as for the researcher (214%)

Several bibliometric indices include peer review as a parameter within their mandate, by according credibility to journals and publications on the basis of their peer review standards. Therefore, even in situations where peer review isn't used as an independent metric, it is understood to be fairly central to the process of establishing credibility around a piece of research. The qualitative, and thus somewhat subjective and diverse, nature of peer review evaluations has led to attempts to create a more standardised way of measuring them. The Peer Review Evaluation Score (pre-SCORE) is a quantitative index that weighs evaluations from multi-round peer reviews, weighing each subsequent round using the square root of earlier rounds in order to account for later rounds often being less rigorous (Etkin 2013). 2. Limitations

B. Limitations

Frequently scrutinised by scholars for their inability to provide standardised measures, peer reviews operate within the following limitations:

- **Academic subjectivity and biases that may emerge during the peer review process are difficult to account for.** The potential impact of factors such as reviewers' specialisations and research interests, and the possibility that reviewers and researchers are unfamiliar with each other's linguistic, socioeconomic, and other contexts, is not accounted for when peer reviews are used as an absolute marker of research quality. Even when there are prescribed parameters for the review process, they are typically put forth by journal editors, and thus similarly subject to their discretion and dispositions. One could argue that qualitative measures often rely on trading off a degree of standardisation in favour of more context and detail. However, there still remains a possibility that the actual outcomes of peer reviews that are meant to denote the quality of research could sometimes end up being arbitrary based on these factors, and hence are also less credible.
- **By design, peer reviews show an over-reliance on in-group perspectives.** Since they are conducted by experts within the same disciplines as researchers, peer review measures run the risk of foregrounding research outcomes that are prioritised within the particular discipline, by those who already hold academic credibility within them. This indicates that they risk not taking into account parameters that focus on the broader impact of research, including how it fares in terms of economic utility, as well as gatekeeping spaces of mainstream research from underserved regions and groups.
- **Delays and prolonged timelines around peer review processes can act as impediments to time-sensitive research, or timelines around large-scale research projects.** Due to its multi-stakeholder nature, peer review can be a time-consuming process, and therefore prevent topical, relevant research from being published swiftly, or from the next steps in larger research projects being carried out in a timely fashion. It is important to note that this limitation does not point to potential inaccuracies in the capacity that peer reviews have to measure research productivity and quality, but instead to logistical inefficiencies caused by their use that can impede research processes as a whole.

MEASURING EQUITY WITHIN RESEARCH

Measures of how equitable research processes and outcomes are, in terms of how many researchers from underserved social groups are able to participate in research in a given unit of research production (such as a discipline or a region), are increasingly being brought into focus within research assessment. One prominent strand of measurement here is that of gender gaps in research productivity. A number of studies today seek to use measures such as bibliometrics and open access indicators to compare differences in how men and women researchers fare while engaging in scientific research.

Figure 4: Measuring Equity within Research

1.4 Economic Productivity-Oriented Measures

A. Definitions

This subsection focuses on research productivity metrics built around measures of economic productivity. There are two categories of metrics that are addressed here. First, we examine indicators that apply economic methods of calculating productivity to research processes, wherein research is treated as an economic activity. Next, we look at indicators that identify the productivity of research as being a function of its economic output, and hence measure the economic impact of research processes. The difference between these two types of indicators is that, while one approach is concerned with using particular economic methods, the other is concerned with prioritising specific economic outcomes.

The first category of metrics identifies inputs, outputs, and other variables within the research process, and uses economic formulae that utilise them to calculate the productivity of research. Economic productivity is defined in general terms as "a ratio of a volume measure of output to a volume measure of input use" (OECD 2001a). Several research productivity metrics have been framed around research being evaluated in similar terms. An OECD (Organisation for Economic Co-operation and Development) economic study from 2001 that analyses R&D and productivity in 16 countries, cites economic tools that regress the multi-factor productivity of countries on stocks of R&D (OECD 2001b). Giovanni Abramo and Ciriaco Andrea D'Angelo (2014) point out that the total factor productivity of research, which requires information on different productive factors such as scientific instruments, time allocations between different types of research (basic vs applied), or researchers' inclinations towards research output in forms outside of publications, involves disparate data that is difficult to standardise. Instead, they propose an indicator called Fractional Scientific Strength (FSS) based on simple labour productivity, ie. the output per unit value of labour within research, that measures the productivity of research in different fields (ibid.). For research productivity rankings that concern non-homogenous units producing research for particular fields, this indicator is further developed such that it measures the productivity of individual researchers or subfields (ibid.). It is important to clarify here that these metrics are not directly focused on measuring the role of research productivity in economic growth. Instead, they focus on treating research as they would any other economic activity, and expressing it in microeconomic terms based on this.

The second category of metrics selects parameters linking research to economic growth, and builds indicators that measure how effectively it is able to do so. The relationship between research productivity and economic output has often been traced through patents. Through various indicators based on the quantity and nature of patents filed in given regions, timeframes, or fields, assessors seek to isolate the economic yield of research activity. Lanjouw and Schankerman (2004) note that while looking at research productivity through patents, it is important to account for the technological dimensions of an innovation, in addition to its value dimensions. Taking this into account, they propose a composite quality index that measures changes in patent quality, correlating them with R&D (ibid.). Other indicators that are often referenced while demarcating research productivity in this manner include number of patents files, types of patents filed, and number of claims under a single patent.

2. Limitations

While economic productivity-oriented assessment metrics would likely be able to yield standardised data around research productivity, they face the following limitations:

- **By design, economic metrics operate on certain assumptions about research activity.** Variables such as the resources available to research units within the same field, and the number of hours devoted to research by each individual, are often not supported by enough relevant data, and hence assumed to be homogenous (Abramo and D'Angelo 2014). These assumptions prioritise a macroscopic view of research productivity, and in the process, concede on details about the functioning of individual research units that could illuminate important patterns and cues around research processes.
- **While measuring research productivity through such metrics often utilises quantitative and bibliometric data, it is unable to accommodate more qualitative metrics.** Given that research output is thus measured in terms of its output quantity by these metrics, they leave significant gaps in measurement of the relevance, quality, and impact of research beyond its publication. When adjusted for quality, they suffer prominently from the problem around assumptions mentioned above.
- **Measuring research productivity in terms of its industrial output and patents filed on its basis can both inaccurately evaluate and significantly discourage certain types of research.** If economic output were to be treated by researchers and evaluators as the dominant marker of research success, time-intensive or high risk research efforts that are unlikely to yield positive short-term economic outcomes would likely be deprioritised.

Section 02

A Review of International Frameworks for Measuring Research Productivity

While one can find an increasing number of guidelines and frameworks that outline composite approaches to research productivity measures, there is no global standard for their usage that has thus far managed to gain widespread acceptance. Arguably, a lack of rigid standardisation lends a flexibility to these measures that allows for metrics to be selected, designed, and deployed in context-sensitive ways, yielding more specific and relevant results. At the same time, the lack of standardised, aggregated research productivity data impedes comparisons between institutions, regions, and countries.

In recent years, several frameworks that attempt to bridge this gap have been brought forth by different bodies. Notably, these frameworks tend to converge on the idea that there is a need to either update traditional measures of research productivity, or to supplement them with other forms of processes and data. This section reviews three such frameworks that have sought to create international standards for measuring research productivity, and examines the gaps in the scope of their application.

2.1 The Leiden Manifesto

The Leiden Manifesto is named after a 2014 STI conference held in the Netherlands, in collaboration with the European Network of Indicator Developers (ENID). It sets out ten principles that denote best practices for metrics-based research assessments (Hicks et al. 2015):

- Quantitative metrics ought to be used with qualitative, expert assessment.
- Research performance should be measured against the research goals of institutions, groups, or researchers.
- Areas of excellence within locally relevant research should be protected.
- Processes around data collection and analysis should be open, transparent, and simple.
- Researchers being evaluated should be allowed to verify evaluation data and analysis.
- Variations by field in publication and citation practices need to be accounted for.
- The assessment of individual researchers should be based on a qualitative judgement of their portfolio, as opposed to purely bibliometric markers.
- Attributing concreteness and false precision to indicators should be avoided.
- The systemic effects of establishing assessment standards and indicators should be recognised, in terms of the incentives they create.
- Indicators should be scrutinised and regularly updated.

The scope of the Leiden Manifesto is somewhat limited, with its best practices guidelines being applicable largely to bibliometric measurement exercises. Its core emphasis relates to why there is a need to move from purely quantitative metrics, towards identifying where and how qualitative indicators can help make them more precise and holistic.

2.2 DORA (The Declaration on Research Assessment)

The Declaration on Research Assessment (DORA) was formed at a 2012 annual meeting of the American Society for Cell Biology in San Francisco, and has since grown into a collaborative global initiative that brings together funders, publishers, institutions, and researchers (DORA, n.d.). DORA identifies a four-step objectives programme— (1) raising awareness, (2) facilitating implementation, (3) catalysing change, and (4) improving equity (ibid.).

The Declaration puts together recommendations on carrying out research assessments for funding agencies, institutions, publishers, organisations supplying metrics, and researchers, based on three prominent themes (ibid.):

- the elimination of the use of journal-based metrics, such as Journal Impact Factors, in processes linked to funding, appointment, and promotion
- the assessment of research on the merits of the research output itself, as opposed to its quantity
- the updation of research quality measurements to capitalise on the new opportunities and areas opened up by online publication

DORA's most recent undertaking, Project TARA (Tools to Advance Research Assessment), works towards facilitating the development of new policies and practices around the criteria used by universities to make hiring, promotion, and tenure decisions. The body's goals primarily revolve around reforming research assessment such that indicators measuring quantitative research output are deprioritised, and there is a renewed focus on evaluating the quality, scientific merit, and accessibility of research output in their place.

2.3 The European Commission Scoping Report 2021

In a scoping report released in November 2021, the European Commission details findings on reforming research assessment processes based on a consultation with European stakeholders. It posits that research assessments ought to serve the function of allowing these stakeholders to evaluate the quality and performance of research in terms of achieving excellence and impact, and building societal trust in the outputs of research and innovation systems (Directorate-General for Research and Innovation 2021). To this end, the report emphasises the need for prioritising ethics and integrity in research, safeguarding the freedom of scientific research, respecting the autonomy of research organisations, and ensuring independence and transparency of the data, infrastructure, and criteria involved in research assessment.

The report envisions a coalition approach through an agreement signed by individual research funding organisations, research performing organisations, and national/regional assessment authorities and agencies (ibid.). It suggests the following principles as the building blocks of such an agreement (ibid.):

- focusing research assessment on quality and transparency
- recognising contributions that advance knowledge and create potential impact
- rewarding behaviours that further open science and open collaboration
- using criteria that are specific to different scientific disciplines, research types, and research career stages
- ensuring gender equality and other forms of equal opportunity in research

While this scoping report puts across a vision for a consensual multi-stakeholder framework, it does not function as an actual framework itself. However, in envisaging a degree of centralisation that encompasses national and regional assessment authorities, it attempts to standardise research assessment processes to a degree that other frameworks have not done so far. In doing so, it consolidates the status of research productivity measurement as an international policy concern.

Section 03

A Thematic Analysis of Gaps in Usage Guidelines

This section analyses the frameworks outlined in the first section, and the metrics reviewed in the second, to identify two thematic areas under research productivity assessment where gaps in usage guidelines pose clear impediments to measurement exercises. We offer recommendations based on these themes to point out key issues that research productivity discourses should tackle in the near future.

3.1 Standardisation gaps and lack of comparative data

One of the recurrent problems that comes up while examining various aspects of research productivity measurement practices is the lack of consensus around when and how particular metrics should be deployed. Some of the global frameworks reviewed earlier in this study aim to create a degree of consensus around these parameters through usage guidelines, but still do not explicitly address how the data yielded from research assessments can be made more standardised. Given the number of diverse quantitative and qualitative tools employed by these measures, as well as the reasonable imperative of making assessments context-specific, it would be fair to conclude that extremely linear and homogeneous research productivity data might in fact act to the detriment of measurement exercises. However, there are compelling reasons for why a degree of standardisation within measurement outcomes is still in order. First, and perhaps most overarching, is the need for research productivity data that can be compared to one another. The availability of comparative data is significantly higher for certain indicators and sample sets, such as university ranking lists, which explicitly assess research productivity with the purpose of comparing and ranking different units. On the other hand, while studies that measure and compare different countries' research productivity have been worked on by transnational bodies such as the OECD, global indices that measure country-wise research productivity using similar metrics are difficult to come by due to the vastly incomparable data that countries tend to produce on the subject. Second, having standardised frameworks of research productivity metrics to reference is key to making measurement exercises more transparent and accountable. Similar to the need for having clearly stated goals ahead of measurement exercises, there is a need to set pre facto expectations and standards that the processes and outcomes of research productivity assessments can be held to. Building frameworks and toolkits that highlight the specific questions that efficiently produced research productivity data should be able to answer thus also holds utility as a tool of verifying the credibility of metrics used and advocating for reforms within them.

Based on these two considerations, we offer the following recommendations when it comes to standardisation:

- The creation of more international guidelines on standardisation would help produce more comparative research productivity data. The international frameworks reviewed in this study do not adequately address how large-scale research assessments at the level of countries or disciplines can be made more streamlined. Establishing standards for this data for large scale assessments at the level of countries or disciplines would make it easier for government agencies to account for research productivity while making STI policies and seeking to improve education outcomes.
- Establishing normative standards around how to use different metrics would help make their usage more accountable as well as accessible. It is pertinent to note that such standards can continue to exist in non-binding forms, alongside an emphasis on context-sensitivity remaining key to their use. In addition to best practices around research productivity measurement exercises at large, best practices within the use of particular metrics would act as a nudge for researchers and practitioners in the domain to adapt to the most holistic and comparative forms in which they could be deployed, and for other stakeholders to ask informed questions about their use as a checking mechanism.

3.2 Context-sensitivity and the need for Global South-oriented metrics

Across global frameworks on research productivity assessment, there has been a move towards making scientific research more equitable and creating indicators that incentivise the same, whether through the growing importance of open science practices or through an emphasis on the development of indices that measure minority participation in research. However, differences in the contexts of developed and developing countries and its implications for research productivity assessment remain an under-analysed subject in global discourse in the area. While studying research productivity measurement in the Pacific Islands to examine how measurement is carried out in a developing country context, Ekeroma, Shulruf, and McCowan (2016) note that the existence of multiple indices and research performance measures in developed countries across the globe has not translated to a prominence in their usage in developing countries. An example of how this plays out can be found in the annual Research and Development Statistics report (2020) released by the Indian government's Department of Science and Technology. The report for 2019-20 focuses on statistical measures around research expenditure, research output data using publication databases including Scopus and SCI (Science Citation Index), and indicators such as number of patents filed by Indian residents. The official country-wide measurement of research, then, does not employ any of the research productivity metrics detailed in this study, besides publication and patent counts.

It is worth noting that the centrality accorded to concerns around economic utility and efficacy of research, as opposed to a detailing of India's qualitative or quantitative research output, makes sense given its status as a developing country. Countries that share this status and are located in the Global South (broadly used to refer to countries outside the

American and European regions) tend to grapple more with resource crunches, infrastructural demands, and relatively new transitions towards knowledge-oriented economies than their Global North counterparts, due to imbalances in global power and differential access to development opportunities historically. This points towards two broad conclusions when it comes to research productivity. First, an entire chunk of the world, encompassing countries from Asia, Africa, and Latin America, does not find adequate representation as far as research productivity measurement metrics and guidelines that take into account their specific needs and constraints are concerned. The call for measurement exercises to be more context-sensitive points towards the need to correct this underrepresentation, but does not engage with developing countries' contexts enough to outline how this ought to pan out. Second, given the imbalances in research development and knowledge production that are endemic to their status as developing countries, research productivity in these countries should in fact be accorded centrality at the global level if equitability is considered a key parameter. Mapping research productivity in the Global South is therefore an area of utmost priority that needs to be addressed.

Taking this urgency into account, we put forth the following recommendations specific to research productivity in developing countries, as well as among researchers facing resource disprivilege:

- **There is an imminent need to create frameworks specific to developing countries, detailing research productivity measures that further shared goals based on economic development.** Rather than dismissing parameters such as expenditure, employment generation, and R&D budgeting, it is pertinent that indicators that serve these needs are looked into and developed to the same degree that other indicators have been over the last few decades. Such an exercise would inevitably involve in-depth engagement with various contexts, needs, and incentives that surround research ecosystems in these countries, and likely be a large-scale undertaking that calls upon researchers from different fields and backgrounds.
- **Measures such as journal impact rankings and citation impact should be made more accommodating of researchers who lack resource privilege.** This entails aspects such as according more credibly to non-English language research output through increasing its inclusion in citation databases and journal repositories, as well as prioritising localised impact while mapping the relevance of research. Not only would this contribute positively to making commonly used research productivity metrics more viable for developing countries to use, it would also make it easier for them to meet goals around standardisation that were outlined in the previous section.
- **Resource constraints that prevent developing countries from accessing particular forms of research productivity success need to be acknowledged and addressed.** Adopting more open access and open science research practices, or increasing gender and other forms of social equity, for instance, are goals that require these countries to invest additional resources when they are already resource-crunched. Measures like publications bearing costs of publication for open access research contributions from developing countries, and the prioritisation of these countries by national and international organisations responsible for funding scientific research, would go a long way in increasing research productivity on these grounds.

Conclusion

Conclusion

Through the course of this report, we have examined various types of literature available in the field of scientific research productivity measurement. While fairly abundant information is available on concepts, frameworks, metrics, and usage guidelines linked to the field, this information is out there in disaggregated forms, sometimes not even being explicitly categorised in terms of research productivity. As such, even though the last decade has seen multiple efforts to build international frameworks on the subject, measurement exercises carried out for various ends across the globe show immense disparity in the methodological tools and approaches they select. We have sought to build common ground using these disparate strands of research productivity literature, and collate resources around it in one place. Prominent approaches to measuring research productivity today warrant several kinds of scrutiny. This report highlights limitations in the applications of commonly used research productivity metrics, as well as international guidelines and frameworks that seek to reform research assessment processes. It identifies two crucial areas that are not satisfactorily addressed by them—the lack of standardised and comparable research productivity data around the world, and the need to develop research productivity parameters that are more inclusive of developing countries' contexts. While these two issues might appear to contradict each other in their scope, our recommendations focus on how accessible and realistic standards can be created to meet the needs created by either of them. These recommendations are summarised below:

- The creation of more international guidelines on standardisation would help produce more comparative research productivity data.
- Establishing normative standards around how to use different metrics would help make their usage more accountable as well as accessible.
- There is also an imminent need to create frameworks specific to developing countries, detailing research productivity measures that further shared goals based on economic development.
- Measures such as journal impact rankings and citation impact should be made more accommodating of researchers who lack resource privilege.
- Resource constraints that prevent developing countries from accessing particular forms of research productivity success need to be acknowledged and addressed.

We conclude this report at a point where more in-depth inquiries can be made into how conversations around scientific research productivity measurement can be made more centralised and accessible, as has been the broad trend in the field. To this end, we hope that it will facilitate the creation of more rigorous and holistic standards for research, and foster a greater focus on STI policies and priorities especially in the developing world, which has thus far been under-represented in this field.

Bibliography

Bibliography

Abramo, G., and C.A. D'Angelo. 2014. "How do you define and measure research productivity?" *Scientometrics* 101, no. 2: 1129-1144. https://doi.org/10.1007/s11192-014-1269-8.

Centre for Civil Society. 2022. "Assessing Scientific Research & Innovation: Study of frameworks and parameters for evaluating institutional research." https://ccs.in/sites/default/files/assessing-scientific-research-_innovation-01.pdf.

Declaration on Research Assessment. n.d. "Read the Decalaration." Accessed on December 17, 2021. https://sfdora.org/read/.

Department of Science and Technology. 2020. R*esearch and Development Statistics 2019-20*. https://dst.gov.in/sites/default/files/Research%20and%20Deveopment%20Statistics%202019-20_0.pdf.

Directorate-General for Research and Innovation (European Commission). 2010. *Assessing Europe's University-Based Research*. https://op.europa.eu/en/publication-detail/-/publication/93ec2eb0-b614-41df-a894-56895a795a54.

—. 2021. T*owards a reform of the research assessment system: Scoping Report*. https://op.europa.eu/en/publication-detail/-/publication/36ebb96c-50c5-11ec-91ac-01aa75ed71a1.

Ekeroma, A.J., B. Shulruf., L. McCowan et al. 2016. "Development and use of a research productivity assessment tool for clinicians in low-resource settings in the Pacific Islands: a Delphi study." *Health Res Policy Sys 14 no. 9*. https://doi.org/10.1186/s12961-016-0077-4.

Elango, B., Dong-Geun Oh, and P. Rajendran. 2021. "Assessment of Scientific Productivity by India and Korea." *DESIDOC Journal of Library & Information Technology 41, no. 3*. https://publications.drdo.gov.in/ojs/index.php/djlit/article/view/16558.

Etkin, Adam. 2013. "A New Method and Metric to Evaluate the Peer Review Process of Scholarly Journals". In *Publishing Research Quarterly 30.* https://link.springer.com/article/10.1007/s12109-013-9339-y.

Hanwell, Marcus. "What is open science?" n.d. Accessed on December 17, 2021. https://opensource.com/resources/open-science.

Hicks, D., P. Wouters, L. Waltman et al. 2015. Bibliometrics: The Leiden Manifesto for research metrics. *Nature* 520: 429–431. https://doi.org/10.1038/520429a.

Kelly, Jacalyn, Tara Sadeghieh, and Khosrow Adeli. 2014. "Peer Review in Scientific Publications: Benefits, Critiques, & A Survival Guide." *EJIFCC 25*, no. 3: 227–243. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4975196/.

Krawczyk, Franciszek, and Emanual Kulczyki. 2021. "How is open access accused of being predatory? The impact of Beall's lists of predatory journals on academic publishing." *The Journal of Academic Librarianship 47, no. 2*. https://doi.org/10.1016/j.acalib.2020.102271.

Lanjouw, Jean O., and Mark Schankerman. 2004. "Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators." *The Economic Journal 114, no. 495*: 441–65. http://www.jstor.org/stable/3590103.

OECD. 2001. Measuring Productivity: OECD Manual: *Measurement of Aggregate and Industry-Level Productivity Growth.* https://www.oecd.org/sdd/productivity-stats/2352458.pdf.

—. 2001. *R&D and Productivity Growth: Panel Data Analysis of 16 OECD Countries*. https://www.oecd.org/economy/growth/1958639.pdf.

—. 2015. "Making Open Science a Reality." *OECD Science, Technology and Industry Policy Papers, No. 25. OECD Publishing: Paris*. https://doi.org/10.1787/5jrs2f963zs1-en.

Piwowar, Heather, Jason Priem, Vincent Larivière et al. 2018. "The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles." *PeerJ 6, no. e4375.* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5815332/.

Priem, J., D. Taraborelli, P. Groth, and C. Neylon. 2010. "Altmetrics: A manifesto." *Altmetrics.* http://altmetrics.org/manifesto.

Suber, Peter. 2004. "A Very Brief Introduction of Open Access." http://legacy.earlham.edu/~peters/fos/brief.htm.

Tercier, John, and Michael L. Callaham. 2007. "Peer Review: Consensus and Contradiction, a Qualitative Approach." *UCSF: Department of Emergency Medicine*. https://escholarship.org/uc/item/2s42m2kw#main.

University of Waterloo Working Group on Bibliometrics. 2016. White Paper: *Measuring Research Outputs Through Bibliometrics*. https://uwaterloo.ca/scholar/s5gordon/publications/university-waterloo-working-group-bibliometrics.

CENTRE FOR CIVIL SOCIETY
A-69 Hauz Khas, New Delhi – 110016
Phone: +91 11 2653 7456 | Website: www.ccs.in