

***“COUNTING WHAT COUNTS: WHY  
BIBLIOMETRICS WIN (AND PEER  
REVIEW LOSES)”***

**Eugene Garfield Centenary Celebration**

**19 Nov. 2025**

***Giovanni Abramo***

*Laboratory for Studies in Research Evaluation  
Mercatorum University, Rome, Italy*

# How Dr Garfield changed the course of my career

*Eugene Garfield transformed me from a technology licensing officer into a research scientist. With an MBA, research was the last thing I ever imagined doing. But everything changed when one of my students in an executive management program brought me a Thomson Scientific Italian National Citation Report CD-ROM filled with Italian bibliographic data. He asked me to be his thesis supervisor — and that's when I had my Eureka moment. With that data, I realized I could evaluate the performance of Italian professors and help curb favoritism in university recruitment. And here I am today, summarizing over 20 years of research in evaluative scientometrics. Thank you, Dr Garfield. I am deeply indebted to you.*

# Research assessment

Research assessment is a generic concept — it may refer to:

- Research products
- Researchers
- Disciplines
- Fields
- Departments/Schools
- Research institutions
- Countries

# Before designing an evaluation exercise ...

- **Objectives:** Identify the objectives of the evaluation.
- **Intended Use:** Determine how results will be utilized.
- **Technical Specifications:**
  - Scale, research disciplines, granularity
  - Timeliness and deadlines
  - Minimum acceptable accuracy level
- **Financial Resources:** Inquire about available budgetary allocations.
- **Data Availability:** Ensure access to necessary data sources.
- **Cultural Context:** Consider societal and cultural factors impacting research and its assessment.

# Peer review vs Bibliometrics

- Peer review and bibliometric methods apply exclusively to the **evaluation of research outputs**.
- How these evaluated outputs are then used to assess other entities depends on **the design of the evaluation exercise**.

# Example of peer review design

- ✓ A = Excellent and extremely relevant (score 1)
- ✓ B = Excellent (score 0.8)
- ✓ C = Standard (score 0.5)
- ✓ D = Sufficient (score 0.2)
- ✓ E = Very poor (score 0)



# Example of bibliometric design

## The Fractional Scientific Strength (FSS)\*

$$FSS_R = \frac{1}{(w_R + k)} \cdot \frac{1}{t} \sum_{i=1}^N c_i f_i$$

Where:

$N$  = number of publications of the researcher in the period under observation

$c_i$  = weighted combination of normalized citations and impact factor associated to publication  $i$ \*

$f_i$  = fractional contribution of the researcher to publication  $i$

$w_R$  = average yearly salary of the researcher

$k$  = average yearly capital used by the researcher

$t$  = number of years of work of the researcher in the period under observation

\* Abramo, G., Aksnes, D.W., & D'Angelo, C.A. (2020). Comparison of research productivity of Italian and Norwegian professors and universities. *Journal of Informetrics*, 14(2), 101023.

# Peer review vs Bibliometrics

In STEMM fields, bibliometrics outperform peer review both in:

- the evaluation of individual research outputs, and
- the assessment of researchers, disciplines, ..., and institutions.



# Bibliometric evaluation limits

- Citation-based metrics cannot be applied to non-indexed research output
- Not all citations are positive or certify real use
- Citations are not representative of all use
- Citations can be gamed: self-citations, network-citations, ...\*

\* Abramo, G., D'Angelo, C.A., & Grilli, L. (2021). The effects of citation-based research evaluation schemes on self-citation behaviour. *Journal of Informetrics*, 15(4), 101204

# Peer review evaluation limits

- Peer review evaluation is highly subjective
- Frequent discrepancies in evaluations by different reviewers
- Peer review systematically penalizes female authored publications more severely than bibliometrics\*
- Peer review quality scores are worse predictors of long-term impact than early citation-based indicators\*\*
- Peer review typically relies on just two referees, whereas the broader scientific community recognizes a work's significance through citations.

\*Abramo, G., & D'Angelo, C.A. (2025). Gender bias in peer review and citation practices: Implications for research evaluation (working paper).

\*\*Abramo, G., D'Angelo, C.A., & Reale, E. (2019). Peer review vs bibliometrics: which method better predicts the scholarly impact of publications? *Scientometrics*, 121(1), 537-554.

# Are reviewers really experts?

## Italian VQR 2015-2019:

- In STEMM, experts reviewed on average 449 works each in 6 months (3.7 works per day).

Question raised:

Was this a genuine peer-review process, as intended, or a flawed scientometric assessment?

- Reviewers are chosen before research works are submitted
- 97.6% of the reviewers were Italian.
- 30% of the reviewers performed below the national median

Question raised: How competent were the reviewers?\*

\* Abramo, G., & D'Angelo, C.A. (2025). Peer review research assessment: Are the reviewers really experts? *Research Evaluation*, 34, rvaf043.

# Peer review sole option

There are specific contexts where **peer review** remains the sole option to assess research performance:

- In arts and humanities (limited coverage of bibliographic databases)
- In national research systems where a significant portion of research works is not indexed in bibliographic repositories

# **Bibliometrics suitability**

In many contexts, scientometrics is more suitable than peer review:

- In STEMM and many social science fields (about 75% of total research)
- In countries where researchers predominantly publish in indexed journals
- As the scale of assessment increases

# Peer review RAEs: main limits\*

- ✓ *Accuracy*: How accurate is the quality evaluation of research outputs and institutions?
- ✓ *Robustness*: How sensitive are rankings to the share of the output evaluated?
- ✓ *Reliability*: Do universities submit their best outputs?
- ✓ *Functionality*: How useful are national rankings for research institutions, companies, stakeholders?
- ✓ *Costs and time of execution*: Implications for frequency

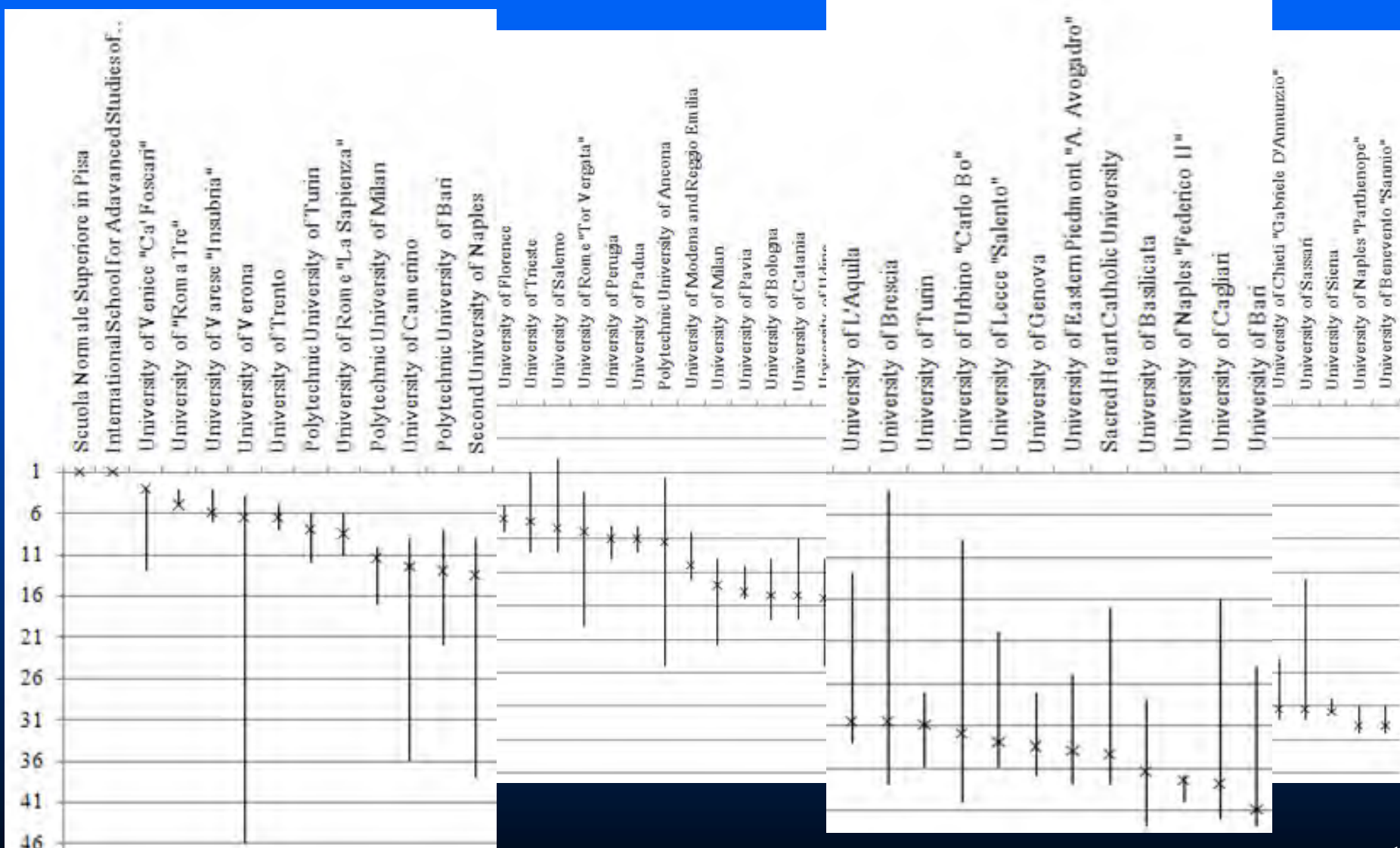
\* Abramo, G. (2024). The forced battle between peer-review and scientometric research assessment: Why the CoARA initiative is unsound. *Research Evaluation*, rvae021



# Peer review RAEs (in)accuracy

- Questionable competence of reviewers
- Time constraints for the evaluation
- The failure to consider outputs' quality scores as a continuous range
- The full counting of the submitted publications regardless of the number of co-authors and their individual contribution to the work
- Limited number of products evaluated (24.4%)
- Inputs (academic rank, gender, ...) not accounted for

# Robustness: Ranking sensitivity to the share of output



Median and range of variation (max - min) of rankings in Physics, when varying output share 8 times

# Reliability: how effective is selection of outputs by universities?

Universities' do-it-yourself selection worsened the maximum score achievable in the STEMM by 23% to 32%, compared to the score from an effective selection.\*

Abramo, G., D'Angelo, C.A., & Di Costa, F. (2014). Inefficiency in selecting products for submission to national research assessment exercises. *Scientometrics*, 98(3), 2069-2086.

# Peer review vs bibliometrics: functionality

Consequences of limiting the number of works to be peer-reviewed:

- Individual level evaluation is not conducted
- Without individual-level assessment, organizations are unable to leverage national assessment outcomes for internal selective funding and individual reward schemes.

# Giovanni Abramo

*Laboratory for Studies in Research Evaluation*  
*Universitas Mercatorum*

[giovanni.abramo@unimercatorum.it](mailto:giovanni.abramo@unimercatorum.it)

<https://sites.google.com/view/laboratoriortt>

# Peer review vs scientometrics: Cost and time of execution

## National Peer-Review Exercises

- Conducted every 5–7 years
- Require 2–3 years from initiation to publication of results
- **Slow and costly**, limiting their usefulness for **timely research management- and policy-making**

## Scientometric Performance Assessments

- Can be **updated frequently** (even daily if needed)
- Completed within **weeks**, not years
- **Far less expensive** (e.g., U.K. REF 2021 cost  $\approx$  £471 million)
- Allow **evaluation of all indexed research outputs**



# Concluding remarks

- The debate between peer review and bibliometrics is too often framed as a battle. It should instead be about balance and intelligent use.
- The community of evaluative scientometricians should reflect on why all RAEs in the industrialized world (but Australia) adopt peer review in STEMM.
- We should finally consider what steps can be taken to ensure that our recommendations for research assessment are effectively implemented.
- Bibliometrics are to research assessment what diagnostic imaging is to medicine. Are CT scans alone enough for a diagnosis? Perhaps not—but could hospitals do without them?