SciencesPo
LABORATOIRE INTERDISCIPLINAIRE
D'ÉVALUATION DES POLITIQUES PUBLIQUES

International Conference on Scientific Impact: Measurement and Policy Evaluation

September 11, 2015, Sciences Po, Paris

Summary Report of Presented Papers

Summary

This conference has been organized by Jérôme Aust, Emeric Henry, and Christine Musselin from the LIEPP- Sciences Po.

The conference aims to question the evaluation of scientific work. Economists, Sociologists, Historians and Political Scientists have participated to the discussion in order to mix the approaches. In the context of Europeans policies of "excellence", how to measure scientific impact? Tools like metrics have been particularly discussed.

Conference Agenda

Session 1:

Clément BOSQUET (University of Cergy-Pontoise, LIEPP-Sciences Po)
"Do large Departments make Academics more Productive? Sorting and Agglomeration Economies in Research"

Dagmar SIMON (Wissenschaftszentrum Berlin für Sozialforschung-WZB)
"Measuring Scientific Activities: Intended and Unintended Consequences... and any Alternatives? Perspectives from the Science Studies"

Session 2:

Alessandro IARIA (CREST, Research Centre in Economics and Statistics-ENSAE)
"International Knowledge Flows: Evidence from the Collapse of International Science in the Wake of WWI", joint paper with Fabian Waldinger (University of Warwick)

Yves GINGRAS (University of Québec in Montréal)
"Criteria for Evaluating Indicators"

Summary of Presentations

Session 1:

Clément BOSQUET (University of Cergy-Pontoise. affiliate London School of Economics, and LIEPP- Sciences Po)

"Do large Departments make Academics more Productive? Sorting and Agglomeration Economies in Research"

Clément Bosquet has presented a collective work joint with Pierre-Philippe Combes (Aix-Marseille School of Economics and Sciences Po), research leader in the study of agglomeration effects. The presentation is based on a work in progress.

Objective:

The project aims at evaluating the impact of academic departments' size on individual researchers' productivity. The structure of academic research is of interest for policy makers: Is it more profitable for the quality of scientific work to agglomerate researchers in large departments or to spread researchers in smaller groups?

Method:

This work borrows methods from the urban economics literature (which analyses the impact of city size and other city characteristics on innovation and productivity) and applies these methods to the field of Economics of Science. The paper controls for the sorting of individual researchers into academic departments and asks whether **there are localisation effects on individual productivity.** In order to evaluate the impact of academic departments' size on research productivity, the identification comes from "movers", academics who moved between large and small departments. They exploit a data set of individual researchers over 19 years to account for selection effects.

Research productivity is measured taking into account the following dimensions:

- the **number of co-authors** per articles: each article is divided by the number of co-authors;
- the **journal quality**: the journals are weighting in function of citation indexes;
- the relative **number of pages** per articles published;
- the **publication delay**: to evaluate the productivity of researchers in the current year the authors choose to use a three years moving average using papers published during the three following years, considering publication delays.

The analysis is based on *Econlit* data, an international data set of the international publications in economics.
The determinants of three outcome variables are considered, the probability to publish, the number of publications, and the average quality of publication.

The following **individual characteristics** of researchers are then considered in the econometric analysis:

- the **position** of the researcher: Is he or she a university professor, an assistant professor, a researcher linked to a research institute as the CNRS? Indeed, positions determine the time that researchers dedicate to research activity.
- the **diversity** of individuals' research activity (researchers publishing in different field or in only one field?).

And the existence of a research consortium/team measured by the average **number of co-authors per paper** and the **connection with co-authors in others countries**.
Selection on unobserved characteristics is controlled in some specification by an individual fixed effect, which consists in estimating the effect on the variations over time of publications for each individual.

The **characteristics of academic departments** are also evaluated trough different variables such as:

- the **composition** of the department (economic and democratic structure of the department, e.g. average age of researchers, share of female researchers);
- the department **size;**
- department diversity of research activity;
- department **heterogeneity** in terms of performances.

As the urban economics literature which studies the effect of market access, this paper tries to capture the **research access** of departments by taking into account the distance between the department studied and the nearing ones.

Results:

Through this study, the authors have succeeded to reveal a role for specialisation and **overall department effects**: the presence of colleagues who belong to a same specific field benefits to the individual productivity of the researcher and it is shown that the location of academics matter for their publication performance However, apart specialisation, it's highly hard to explain the channels of these department externalities on scientific productivity.

Finally, for a given local environment, the authors shed some light on the individual determinants of publications. Nowadays, young researchers have more incentives to publish than 20 years ago, which could explain why older researchers publish less. **New publishing strategies** are emerging, benefiting from increasing returns to co-authorship for quality: co-authors publications permits to reach "top-journals". Related projects using the same date study the effect of gender on the probability of promotions and look for natural experiments to test for directions of causality.

Questions:

Clément Bosquet was questioned about individual quality evaluation measured by the quality of journals in which the researcher publishes articles. It could be more useful to measure the individual quality of scientific work through citations. Yves Gingras insists on the utility of a co-authorship variable to measure new research productivity strategies.

Christine Musselin questioned the method consisting of identifying the effects of departements' characteristics through "movers". Why not choose to study people in department with different characteristics and size but with same individual characteristics at the beginning? How to take into account past trajectory of people before they move? Others point out that moving person can move because of very different reasons.

Alessandro Iaria puts the light on the role of the division of scientific work on quality productivity. But he asked himself on the way to measure this division of scientific labour.

Yves Gingras: Concerning the co-authorship, the nationality of the co-author doesn't weight the same, and could facilitate more or less the access to top journals. The quality is not only an abstract variable, here sociology is helpful.

Christine Musselin quotes a qualitative study lead by Felippe Camerati in the UK insisting on the role of the type of departments: bureaucratic department are less productive than collegial departments.

Dagmar SIMON (research group « Science Policy Studies », Wissenschaftszentrum Berlin für Sozialforschung-WZB)

"Measuring Scientific Activities: Intended and Unintended Consequences... and any Alternatives? Perspectives from the Science Studies"

Objective:

Her work focuses on the way of measuring the scientific work. She concentrates her researches on the use of **bibliometrics** in academic and research fields. Studying the metrics, she aims at opening the "black box" of the production of scientific knowledge.

First, the measure of scientific work deals with a **scale and size problem**. Indeed, the choice of associated variables is very important: what is important? Teams? Departments? Individuals? Universities? Nowadays, the quantitative indicators are used to reduce the complexity. But we observe a **diversification of the quantitative indicators**. Such practices lead to transform the scientific field: dominance of reference journals, modification of publication strategies with impact factors. Metrics engaged **unintended effects**.

Method:

Her work takes from the sociology of science and from the philosophy of sciences at the same time. She based her study on ***The Metric Tide,*** a report lead in Great-Britain by the Independent Review of the Role of Metrics in Research Assessment and Management and published in 2015. It deals with the use of metrics in the scientific field. This report constituted a basis for Dagmar's work. Indeed, as a science policy researcher, Dagmar Simon works on an "**excellence initiative**" focused on the German University system. Between 2006 and 2007, the Universities received 4.6 billion in order to fulfil three funding lines of the initiative:

1. Working on graduate schools to promote early career to researchers
2. Build Clusters of excellence
3. Develop new institutional strategies.

This project of "excellence" reform enter in the actual tendency described by Christine Musselin (2006)[1], turning universities into organizational actors.

This plan of Universities' development leads to transformation of the role of the Universities' presidents. Presidents setting up mega structures (clusters, research council…) create new disciplinary sectors and modify the organization of scientific development.

She develops an expertise based on political science studies as the **institutional isomorphism** from Di Maggio and Powel (1983)[2].

On her presentation she also dealt with the question of the reputation system.

Results:

Metrics are a basis for information and not a **key for evaluation**; they couldn't and should not replace the **expert judgment**.

---

[1] MUSSELIN, C. (2006) "Are Universities specific organisations?" in Krücken G., Kosmützky A. et Torka M. (eds.): *Towards a Multiversity? Universities between Global Trends and national Traditions*, Bielefeld, Transcript Verlag, pp. 63-84.

[2] DI MAGGIO P., POWELL W. (1983) "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields", *American Sociological Review*, volume 48, issue 2, pp.147-160.

Questions:

Emeric Henry points out a solution to this debate: What do we choose between expert evaluation and metrics evaluation? A mixed solution?

Yves Gingras insists on the problem of scale. Metrics indeed are a question of scale: what are we evaluating? Do we evaluate individuals? Researches of a laboratory? […] And he adds that quantitative information on science is different to metrics. He warns against the transparency illusion given by metrics. Dagmar Simon adds that the way of peers use metrics is also an important point to analyse.

Clémentine Gozlan: What are the differences and articulations between metric systems and reputational systems? Dagmar Simon referred to the literature on "reputational system" which is very rich. Reputation system takes into account the structure of scientific community.

Christine Musselin: There is also different kind of making peer evaluation. And many kind of this peer review are not analysed yet.

Session 2:

Alessandro IARIA (CREST, Research Centre in Economics and Statistics-ENSAE)

"International Knowledge Flows: Evidence from the Collapse of International Science in the Wake of WWI", joint paper with Fabian Waldinger (University of Warwick)

Objective:

By analysing citations in scientific papers, Alessandro and Fabian investigate whether more costly access to existing knowledge may slow down future production of knowledge. More broadly, the authors try to evaluate the importance of international cooperation for the progress of science.

Method:

Alessandro and Fabian measure international knowledge flows with citations in scientific papers. Cross-country differences in citation patterns, however, do not only reflect differences in access to knowledge but also other forms of cross-country heterogeneity, such as a differential specialization of scientists. To isolate knowledge flows from these other differences, the authors rely on a sudden and sharp change to international knowledge flows that affected the entire scientific community in the wake of World War I (WWI): the so called "boycott."

The boycott interrupted knowledge flows between Allied and Central nations. Scientists from Allied countries suddenly faced higher barriers to access knowledge from Central countries; in particular from Germany, a country whose scientists had received more than 40 percent of Nobel prizes in physics and chemistry in the pre-war period. Similarly, scientists from Central countries faced higher barriers to access knowledge from Allied countries; in particular from the UK (20 percent of Nobel prizes), France (15 percent of Nobel prizes), and the rising scientific power United States.

Alessandro and Fabian estimate how much WWI and the boycott reduced international knowledge flows by analyzing citation patterns in five scientific fields: medicine, biology, chemistry, physics, and mathematics. Readily available publication and citation data lack address information for authors and cited references for the historical period studied. As country information is essential to study international citation flows, the authors construct a new dataset of all university scientists in the world in 1900 and 1914. These data contain names, scientific spe-

cializations, universities, and thus country affiliations for all university scientists. Alessandro and Fabian combine the scientist data with data on more than 260,000 articles citing almost 2 million references from over 150 journals from the ISI Web of Science.

Conclusions:

The boycott increased citation penalties against enemy countries by around 100%, indicating a substantial reduction in international knowledge flows. Additional results show that the authors' findings are not driven by discrimination against enemy papers but rather by a genuine reduction in knowledge flows, and that some knowledge that was produced during the boycott never reached the enemy camp. Alessandro and Fabian also provide suggestive evidence that the collapse of international science affected the world-wide production of Nobel Prize worthy ideas.

Questions:

Emeric Henry: How can we take into account the scientific migration because of the expulsion of the Jews from Nazi Germany?

In such period there was international cooperation and migration. Jewish scientists went to the US and so the probability that central papers quote allied citations increased after 1933.

Christine Musselin notices that it could exist before WW1 some kind of link between countries that favour international cooperation. How can we take into account these cooperation links before WWI?

Yves GINGRAS (University of Québec in Montréal, Canada research Chair in the History and Sociology of Science)

"Criteria for Evaluating Indicators"

Objective:

During the last ten years, an excitation for evaluation emerges, an evaluation based on metrics. Yves Gingras proposes to analyse what he calls an "**Evaluation Fever**". Now, there are **many indicators for different levels**. And scientists improvised themselves as evaluators. We observe the emergence of a **new market of ranking**. But in order to be **valid,** these indicators must have specific properties. The problem is that the users of these indicators never check their validity using. I thus propose to define the criteria that must be obeyed by any **indicator** to assure their validity.

Method:

Despite the large number of indicators used in different rankings of universities, for example, very few are in fact valid. In addition to the criteria of validity we defined, there are other criterias to evaluate the indicators: timeless, behavioural impact, level of aggregation, transparency and relevance. In 2006, the Berlin Principles for Ranking proposed the following criteria for indicators: a transparent methodology, choosing indicators according to relevance and validity, but these criterias are not even met by existing rankings.

Yves Gingras has developed **prerequisites for indicators' validity**:

- The indicator must be in adequation to the concept he is supposed to measure.
- The indicator must take into account the inertia of the phenomenon it measures (here scientific institutions).
- The measure of the indicator must be homogeneous; one should not add different

kinds of units and weigh them in an arbitrary manner.

- The relation between the concept and its indicator must be monotonous: a higher value for the indicator must always mean a higher value of the concept and vice versa.

Many current indicators used in the evaluation of scientific research such as the **h-index** are not valid. There are also other metrics indicators based on web measures (« altmetrics » as twitter) for which nobody really knows what they measure exactly and what they mean. The « altmetrics » are supposed to replace citations whereas they are simply a different measure of a different kind of diffusion of research.

Conclusion:

Indicators of productivity and scientific impact are now widely used in many different **ranking of universities but most of these indicators are not well evaluated to make sure they are valid**. In using valid indicators, it is also necessary to **evaluate the quality of the data base on which the indicators are built.**

Questions:

Christine Musselin: Many people use indicator knowing they are bad. Why are they using them knowing this? Why scientometricians are not heard by the policy makers? Is this a problem of reception?

According to Yves Gingras, policy makers need such indicators to assess research and so they're using them even though they are wrong; for they care more about showing that they do make evaluations than about the validity of what they are doing. In short: showing that they **do** something is more important for them than making sure that what they do really make sense.