



# Quantitative science studies should be framed with middle-range theories and concepts from the social sciences

Thomas Heinze<sup>1</sup>  and Arlette Jappe<sup>2</sup> 

<sup>1</sup>Institute of Sociology, University of Wuppertal, Germany

<sup>2</sup>Interdisciplinary Center for Science and Technology Studies (IZWT), University of Wuppertal, Germany

an open access  journal



Citation: Heinze, T., & Jappe, A. (2020). Quantitative science studies should be framed with middle-range theories and concepts from the social sciences. *Quantitative Science Studies*, 1(3), 983–992. [https://doi.org/10.1162/qss\\_a\\_00059](https://doi.org/10.1162/qss_a_00059)

DOI:  
[https://doi.org/10.1162/qss\\_a\\_00059](https://doi.org/10.1162/qss_a_00059)

Corresponding Author:  
Thomas Heinze  
[theinze@uni-wuppertal.de](mailto:theinze@uni-wuppertal.de)

Handling Editors:  
Loet Leydesdorff, Ismael Rafols,  
and Staša Milojević

Copyright: © 2020 Thomas Heinze and Arlette Jappe. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.



**Keywords:** bibliometrics, Italy, middle-range theory, Netherlands, profession, research assessment

## ABSTRACT

This paper argues that quantitative science studies should frame their data and analyses with middle-range sociological theories and concepts. We illustrate this argument with reference to the “sociology of professions,” a middle-range theoretical framework developed by Chicago sociologist Andrew Abbott. Using this framework, we counter the claim that the use of bibliometric indicators in research assessment is pervasive in all advanced economies. Rather, our comparison between the Netherlands and Italy reveals major differences in the national design of bibliometric research assessment: The Netherlands follows a model of bibliometric professionalism, whereas Italy follows a centralized bureaucratic model that co-opts academic elites. We conclude that applying the sociology of professions framework to a broader set of countries would be worthwhile, allowing the emerging bibliometric profession to be charted in a comprehensive, and preferably quantitative, fashion. We also briefly discuss other sociological middle-range concepts that could potentially guide empirical analyses in quantitative science studies.

## 1. INTRODUCTION

The argument of this paper is that quantitative science studies should more frequently frame their data and analyses with middle-range sociological theories and concepts (Merton, 1968) in order to advance our understanding of institutional configurations of national research systems and their changes. We illustrate this argument with reference to the theoretical framework “sociology of professions” (Abbott, 1988, 1991), which we apply to a comparison of research evaluation frameworks in the Netherlands and Italy, two countries that contrast in how knowledge for research assessment is both produced and used.

We argue that, as a new and emerging profession, evaluative bibliometrics has successfully established a subordinate jurisdiction in the Netherlands, but that similar professionalization cannot be observed in Italy. Our comparison of these two countries suggests that the institutionalization of bibliometrics via expert organizations in the context of decentralized decision making among universities, such as in the Netherlands, generates trust and learning, contributing to increased scientific performance (so-called improvement use of research evaluation, cf. Molas-Gallart, 2012). In contrast, if research assessments are institutionalized via centralized government bureaucracies with little or no involvement of bibliometric expertise, such as

in Italy, it generates little trust and learning, and tends to be an instrument of administrative control (so-called controlling use of research evaluation, cf. Molas-Gallart, 2012).

Our starting point is the reportedly increased use of bibliometric indicators in research assessment (Hicks, Wouters, et al., 2015; Wilsdon et al., 2015). Such metrics are commonly based on both publication and citation data extracted from large multidisciplinary citation databases, most importantly the Web of Science (WoS) and Scopus. The simplest and most common metrics include the Journal Impact Factor and the Hirsch Index (Moed, 2017; Todeschini & Baccini, 2016). An influential narrative as to why such metrics have proliferated is increased accountability pressures in the governance of public services, including research and higher education, reflecting a global trend towards an audit society in which public activities come under ever-increasing scrutiny (e.g., “governance by numbers” or “metric tide”) (Espeland & Stevens, 1998; Miller, 2001; Porter, 1995; Power, 1997; Rottenburg, Merry, et al., 2015). With regard to research assessments, these metrics have led to considerable criticism, not only in methodical terms (Adler, Ewing, & Taylor, 2009; Cagan, 2013) but also with respect to potential negative side effects, such as the suppression of interdisciplinary research or the sub-optimal allocation of research funding (van Eck, Waltman, et al., 2013; Wilsdon et al., 2015).

Some believe that research assessment metrics are used pervasively in all advanced economies, regardless of the national institutional context in which the research is carried out. The contestation of such metrics among scientific stakeholders is said to be indicative of their pervasiveness. However, as we show here, such a view receives little support in light of empirical evidence about how national research systems have institutionalized the professional use of such metrics. Our country comparison clearly shows differences in the design of the Dutch and Italian research evaluation frameworks, and a sociology of professions framework contributes to analyzing these differences (Abbott, 1988, 1991). The Netherlands and Italy differ considerably in their institutional setup, providing very different contexts for the professional activities of bibliometric experts.

First, we conclude that it would be worthwhile to apply the middle-range sociology of professions framework to a broader set of countries. In doing so, the strength of the emerging bibliometric profession could be charted in a comprehensive, and preferably quantitative-descriptive, manner. Second, we point out that insights into the emerging bibliometric profession should be combined with other important institutional factors, such as organizational autonomy in higher education systems. We suggest that cross-national performance comparisons should make a greater effort to include such theoretically framed explanatory variables in multivariate models. Third, we argue that many other sociological middle-range theories and concepts have potential for guiding empirical analyses in quantitative science studies. We briefly discuss one such concept, Hollingsworth’s (2004, 2006, p. 425–426) “weakly versus strongly regulated institutional environments.”

## 2. THE SOCIOLOGY OF PROFESSIONS FRAMEWORK

The work of professionals in modern societies has been described as the application of abstract knowledge to complex individual cases. The application of such knowledge includes diagnosis, inference, and treatment, and is typically carried out in particular workplaces, such as hospitals or professional service firms. Abbott (1988, p. 35–58) identified three social arenas in which professionals must establish and defend their jurisdictional claims: the legal system, the public sphere, and the workplace. From a system perspective, professional groups compete for recognition of their expertise and seek to establish exclusive domains of competence (“jurisdictions”). Abbott argued that historical case studies should be conducted to better understand

the variety of jurisdictional settlements in modern societies. One such case study is the field of evaluative bibliometrics, in which two main types of clients are interested in bibliometric assessment services: organizations conducting research and funders of research (Jappe, Pithan, & Heinze, 2018). Such organizations would be interested in quantitative assessment techniques because of the rapid growth of science (Bornmann & Mutz, 2015). As the knowledge base in most areas of science grows faster globally than the financial resources of any individual organization, these organizations routinely face problems with both resource allocation and staff recruitment. Bibliometric expertise is provided mostly by individual scholars, contract research organizations specializing in assessment services, or database providers offering software for ready-made bibliometric tools, as well as more customized assessment services.

A recent study investigated bibliometric experts in Europe (Jappe, 2019). Based on a comprehensive collection of evaluation reports, expert organizations, such as the Dutch Centre for Science and Technology Studies (CWTS) in Leiden, are able to set technical standards with respect to data quality by investing in in-house databases with improved WoS data. Bibliometric indicators based on field averages were most frequently used. Importantly, the study found that bibliometric research assessment occurred most often in the Netherlands, the Nordic countries, and Italy, confirming studies focusing on performance-based funding systems (Aagaard, Bloch, & Schneider, 2015; Hicks, 2012; Sandstrom & Van den Besselaar, 2018). Yet, how successful have bibliometric experts been at establishing what Abbott (1988: 35–58) calls “professional jurisdictions”?

### 3. RESEARCH EVALUATION FRAMEWORKS IN THE NETHERLANDS AND ITALY

The Netherlands has a tradition of decentralized research evaluation (Van Der Meulen, 2010; van Steen & Eijffinger, 1998). The Dutch evaluation framework is based on the principles of university autonomy, leadership at the level of universities and faculties, and accountability in research quality. In contrast, Italy has developed a highly centralized research evaluation exercise. The Italian evaluation framework is based on financial rewards for publication and citation performance, provides national rankings of university departments, and contributes to an institutional environment that leaves little room for university autonomy (Capano, 2018).

#### 3.1. The Dutch Research Evaluation Framework

In the Netherlands, the Standard Evaluation Protocol (SEP) regulates institutional evaluation (i.e., evaluation of research units at universities), including university medical centers, and at the institutes affiliated with the Netherlands Organization for Scientific Research (NWO) and the Royal Netherlands Academy of Arts and Sciences (KNAW). Three consecutive periods of the SEP have been implemented thus far, from 2003–2009, 2009–2015, and 2015–2021 (VSNU, KNAW, & NWO, 2003, 2009, 2014). The responsibility for formulating the protocol lies with the KNAW, NWO, and the Association of Universities in the Netherlands (VSNU). The legal basis is the Higher Education and Research Act (WHW), which requires regular assessment of the quality of activities at universities and public research institutions.

Research evaluation under the SEP is decentralized, in that evaluations are commissioned by the boards of individual research organizations. No national agency is tasked to bring together the information from different institutions or exercise central oversight over the evaluation process (van Drooge, Jong, et al., 2013). The aim of the SEP is to provide common guidelines for the evaluation and improvement of research and research policy based on expert assessments. The protocol requires that all research be evaluated once every 6 years. An internal midterm evaluation 3 years later serves to monitor measures taken in response to the

external evaluation. The external evaluation of scientific research applies to two levels: the research institute as a whole and its research programs. Three main tasks of the research institute and its research programs are assessed: the production of results relevant to the scientific community, the production of results relevant to society, and the training of PhD students. Four main criteria were considered in the assessments conducted thus far: quality, productivity, societal relevance and vitality, and feasibility. The goals of the SEP are to improve the quality of research, to provide accountability for the use of public money towards the research organization's board, funding bodies, government, and society at large (van Drooge et al., 2013). During the most recent period, the productivity criterion has been abandoned in favor of greater emphasis on societal relevance (Petersohn & Heinze, 2018).

A precursor to the SEP was the VSNU protocol, which was developed in the early 1990s by VSNU in consultation with the NWO and KNAW. In contrast to the current protocol, the VSNU protocol was designed as a national disciplinary evaluation across research organizations. In most disciplines, research quality was assessed by a combination of peer review and bibliometric data. In response to criticism, this assessment framework was overhauled in 1999–2000 and the national comparison of academic disciplines was abandoned in favor of greater freedom for universities to choose the format in which they wanted to conduct their research quality assessment while maintaining a common procedural framework. The responsibility for commissioning evaluations was moved from the disciplinary chambers to the executive boards of research organizations (Petersohn & Heinze, 2018).

### 3.2. The Italian Research Evaluation Framework

In Italy, the Evaluation of Research Quality (VQR) is a national evaluation exercise implemented by the National Agency for the Evaluation of Universities and Research Organizations (ANVUR). The evaluation is mandatory for all public and private universities, as well as 12 national research organizations funded by the Ministry of Education, Universities, and Research (MIUR), involving all researchers with fixed-term or permanent contracts. The current legal basis is law no. 232/2016, which requires that the VQR be carried out every 5 years on the basis of a ministerial decree. The VQR has been completed for the periods 2004–2010 (VQR I) and 2011–2014 (VQR II), and will be continued in 5-year periods, with the next in 2015–2019 (VQR III) (ANVUR, 2013, 2017). The periods refer to the publication years of research output assessed in the respective cycle.

The objective of the VQR is to promote improvement in the research quality of the assessed institutions and to allocate the merit-based share of university base funding. Performance-based funding is implemented as an additional incentive for institutions to produce high quality research. Law 98/2013 dictates that the share of annual base funding (Fondo di Finanziamento Ordinario) distributed to these organizations as a premium (i.e., in large part dependent on their VQR results) will increase annually up to a level of 30%, reaching 23% in 2018. The assessment builds on the administrative division of the Italian university system into 14 disciplinary areas and produces national rankings for university departments within each area based on several composite indicators of research quality. Evaluations are based on the submission of a fixed number of research products per employed researcher. For example, VQR I required three products per university researcher, and six products for researchers working in a nonuniversity setting without teaching obligations. The university or research institute collects the submitted products and selects the final set submitted to ANVUR. These outputs are then assigned to one of the 14 (VQR I) or 16 (VQR II) groups of evaluation experts (GEVs). Research quality is judged in terms of scientific relevance, originality and innovation, and

internationalization. The GEVs are assigned to rate each individual product on a five-point scale, relying on bibliometric information or peer review or a combination of the two. VQR I involved nearly 185,000 research products for 61,822 researchers (Ancaiani, Anfossi, et al., 2015). ANVUR then computes a set of composite quality indicators for each department, university, and research institute (Ancaiani et al., 2015).

A precursor to the VQR was the Triannual Research Evaluation (VTR), which was performed in 2004–2006 with reference to the period 2001–2003. The VTR was inspired by the Research Assessment Exercise in the United Kingdom (RAE); it was an expert review organized into 20 panels to assess the quality of submissions from researchers in all Italian universities and research organizations (Geuna & Piolatto, 2016). The VTR reassessed approximately 14% of the research produced by the Italian academic system during the respective period, relying exclusively on peer review (Abramo, D'Angelo, & Caprasecca, 2009). Its results affected university funding to a very limited extent.

#### 4. PROFESSIONALIZATION OF BIBLIOMETRIC EXPERTISE IN THE NETHERLANDS AND ITALY

The differences in the design of the Dutch and Italian research evaluation frameworks are related to the question of how bibliometric expertise has been institutionalized in each country. From a theoretical point of view, research assessments can be understood as an intrusion upon reputational control that operates within intellectual fields (Whitley, 2000, 2007). However, research assessments do not replace reputational control, but are an additional institutional layer of work control. The sociological question is how this new institutional layer operates.

The Dutch system follows a model of bibliometric professionalism. In the Netherlands, there is a client relationship between universities and research institutes, with a legally enforced demand for regular performance evaluation on one side and primarily one contract research organization, the CWTS, on the other, providing bibliometric assessment as a professional service. In a study on the jurisdiction of bibliometrics, Petersohn and Heinze (2018) investigated the history of the CWTS, which developed as an expert organization in the context of Dutch science and higher education policies beginning in the 1970s. Even though bibliometrics are not used by all Dutch research organizations or for all disciplines, and some potential clients are satisfied with internet-based “ready-made indicators,” the current SEP sustains a continuous demand for bibliometric assessment. In the 2000s, the CWTS became an established provider, exporting assessment services to clients across several European countries and more widely influencing methodological practices within the field of bibliometric experts (Jappe, 2019).

Regarding professional autonomy, there are two important points to consider in the Dutch system. First, the additional layer of control via the SEP is introduced at the level of the research organization, in Whitley's (2000) terms, at the level of the employing organization, rather than the intellectual field or discipline. Thus, the purpose of research evaluation is to inform the university or institute leadership about the organization's strengths and weaknesses regarding research performance. It is the organization's board that determines its information needs and commissions evaluations. In this way, the role of the employing organization is strengthened vis-à-vis scientific elites from the intellectual field, as the organizational leadership obtains relatively objective information on performance that can be understood and used by nonexperts in their respective fields. However, this enhancement of work control seems to depend on a high level of acceptance of the bibliometric services by Dutch academic elites as professionally sound and nonpartisan information gathering. As Petersohn and Heinze (2018)



emphasized, professional bibliometricians in the Netherlands have claimed a jurisdiction that is subordinate to peer review.

This leads to the second point, which is related to the professional autonomy of bibliometrics as a field of experts. In the Dutch system, science and higher education policies have helped create and sustain an academic community of bibliometricians in addition to the expert organization (i.e., the CWTS). The development of bibliometric methods is left to these bibliometric professionals. Neither state agencies nor the universities, as employing organizations, claim expertise in bibliometric methodology. On the other hand, for a professional organization to gain social acceptance of its claims of competence, the CWTS is obliged to closely interact with its clients in order to determine the best way to serve their information needs. Thus, the professional model of bibliometric assessment in the Netherlands strengthens the leadership of employing organizations and supports the development of a subordinate professional jurisdiction of bibliometrics with a certain degree of scientific autonomy. The model of bibliometric professionalism seems to have contributed to the comparatively broad acceptance of quantitative performance evaluation in the Dutch scientific community.

In stark contrast, the Italian system follows a centralized bureaucratic model that co-opts academic elites. In Italy, bibliometric assessment is part of a central state program implemented in 14 disciplinary divisions of public research employment. Reputational control of academic work is taken into account insofar as evaluation is carried out by members of a committee representing disciplinary macroareas. It is the responsibility of these evaluation committees to determine the details of the bibliometric methodology and evaluation criteria appropriate for their area of discipline, whereas ANVUR, as the central agency, specifies a common methodological approach to be followed by all disciplines (Anfossi, Ciolfi, et al., 2016). In this way, the Italian state cooperates with elites from intellectual fields in order to produce the information required for performance comparisons within and across fields. VQR I comprised 14 committees with 450 professorial experts; VQR II comprised 16 committees with 436 professorial experts.

When comparing the two systems, two points are notable regarding the development of a new professional field. First, in the Italian system, an additional layer of work control was introduced at the level of a state agency that determines faculty rankings and at the MIUR, which is responsible for the subsequent budget allocation. Arguably, the role of organizational leadership at universities and research institutes is not strengthened, but circumscribed, by this centralized evaluation program. All public research organizations are assessed against the same performance criteria and left with the same choices to improve performance. Administrative, macrodisciplinary divisions are prescribed as the unitary reference for organizational research performance. Furthermore, although the VQR provides rectors and deans with aggregated information concerning the national ranking positions of their university departments, access to individual performance data is limited to the respective scientists. Thus, the VQR is not designed to inform leadership about the strengths of individuals and groups within their organization. This could be seen as limiting the usefulness of the VQR from a leadership perspective. In addition, the lack of transparency could give rise to concerns about the fairness of individual performance assessments. There seems to be no provision for ex-post validity checks or bottom-up complaints on the part of the research organization or at the individual level. This underlines the top-down, central planning logic of the VQR exercise.

The second point relates to the professionalism of bibliometrics. Italy has expert organizations with bibliometric competence, such as the Laboratory for Studies in Research Evaluation (REV lab) at the Institute for System Analysis and Computer Science of the Italian Research

Council (IASI-CNR) in Rome. However, the design and implementation of the national evaluation exercise has remained outside their purview. For example, REV lab took a very critical stance with regard to the VTR and VQR I. Abramo et al. (2009) criticized the fact that many Italian research organizations failed to select their best research products for submission to the VTR, as judged by an ex-post bibliometric comparison of submitted and nonsubmitted products. Accordingly, the validity of the VTR was seriously questioned, as their conclusion suggests: “the overall result of the evaluation exercise is in part distorted by an ineffective initial selection, hampering the capacities of the evaluation to present the true level of scientific quality of the institutions” (Abramo et al., 2009, p. 212).

The VQR has been criticized by the same authors for evaluating institutional performance on the basis of a partial product sample that does not represent the total institutional productivity and covers different fields to different degrees (Abramo & D’Angelo, 2015). The combination of citation count and journal impact developed by ANVUR was also criticized as being methodically flawed (Abramo & D’Angelo, 2016). This criticism is further substantiated by the methodological design developed by ANVUR for bibliometric-based product ratings clearly deviating from the more common approaches in bibliometric evaluation practice in Europe (Jappe, 2019). Reportedly, the VTR/VQR evaluation framework was introduced by the state against strong resistance from Italian university professors (Geuna & Piolatto, 2016). As shown by Bonaccorsi (2018), the involved experts have exerted great effort to build acceptance of quantitative performance assessment among scientific communities outside the natural and engineering fields.

In summary, the centralized model of bibliometric assessment in Italy severely limits university autonomy by directly linking centralized, state-organized performance assessment and base funding allocation. Although the autonomy of reputational organizations is respected in the sense that intellectual elites are co-opted into groups of evaluating experts, evaluative bibliometricians are not involved as independent experts. In contrast to the situation in the Netherlands, the current Italian research evaluation framework has not led to the development of a professional jurisdiction of bibliometrics.

## 5. FUTURE RESEARCH AGENDA

What could be fruitful avenues for future research? First, we think it would be worthwhile to apply the analytical framework of the sociology of professions to a broader set of countries. Similar to the Netherlands–Italy comparison, such analyses could ascertain the extent to which professional jurisdictions have been established in countries where publication- or citation-based metrics are regularly used in institutional evaluation, including Australia, Denmark, Belgium (Flanders), Finland, Norway, Poland, Slovakia, and Sweden. These findings could then be contrasted with countries that do not operate such regular bibliometric assessments at the institutional level, including France, Germany, Spain, and the United Kingdom. Based on current knowledge (Aagaard et al., 2015; Hicks, 2012; Kulczycki, 2017; Molas-Gallart, 2012; Petersohn, 2016), it is reasonable to assume that countries performing regular bibliometric assessments with the help of recognized expert organizations have developed similar jurisdictions (i.e., subordinate to peer review) as in the Netherlands. Possible examples would be the Center for Research & Development Monitoring (ECOOM) in Flanders (Belgium) or the Nordic Institute for Studies in Innovation, Research, and Education (NIFU) in Norway. In countries without such regular bibliometric assessments, we would expect co-optation of scientific elites into state-sponsored evaluation agencies. Possible examples would be the National Commission for the Evaluation of Research Activity (CNEAI) in Spain and the Science

Council (WR) in Germany. Ultimately, these analyses could chart the institutional strength of the emerging bibliometric profession in Europe, and globally, in a comprehensive and preferably quantitative-descriptive manner.

Second, these theoretically framed insights on the emerging bibliometric profession could be juxtaposed with other institutional dimensions that are important for the scientific performance of national research systems. One such dimension, according to Hollingsworth (2006), is the autonomy of universities to recruit senior academic staff and decide on their promotion, salaries, and dismissal. In this regard, the autonomy scoreboard provided by the European University Association (Pruvot & Estermann, 2017) shows that Dutch universities have higher “staffing autonomy” scores (73 out of 100) than Italy (44). Furthermore, the most recent Science & Engineering Report (NSB, 2019) indicates that the Netherlands had a higher and increasing share of S&E publications in the top 1% of most-cited articles in the Scopus database between 1996 and 2014 than Italy. Does that mean that a country’s impact in science is related to the institutional strength of its bibliometric profession and the staffing autonomy of its universities? We are far from making such a bold claim, because there seems to be no linear relationship between these two points, as exemplified by the United Kingdom (weak bibliometric profession but strong university autonomy). Rather, we suggest that cross-national/regional performance comparisons (Bonaccorsi, Cicero, et al., 2017; Cimini, Zaccaria, & Gabrielli, 2016; Leydesdorff, Wagner, & Bornmann, 2014) should make greater effort to include both explanatory variables in their multivariate models to ascertain whether they are competing and complementary. We would like to reiterate that such variables need to be anchored in middle-range social scientific frameworks, otherwise it will be difficult to build cumulative knowledge. Such variables could be included at various levels of measurement depending on availability and/or data quality: nominal level (dummy variables), ordinal level (categorical/rank variables), or interval/ratio levels (count variables).

Third, although we have discussed the emerging bibliometric profession with reference to the sociology of professions framework thus far, there are clearly other suitable middle-range “candidate theories/concepts” with considerable potential for further quantitative science studies is Hollingsworth’s (2004, 2006: 425–426) concept of “weakly versus strongly regulated institutional environments.” Based on extensive interviews with prize-winning scientists in biomedicine, Hollingsworth argues that universities and research organizations with high numbers of scientific breakthroughs are often found in weakly regulated institutional environments, whereas strong control constrains the capabilities of research organizations to achieve breakthroughs. More specifically, in weakly regulated institutional environments, research organizations have considerable decision-making authority on whether a particular research field will be established and maintained within their boundaries, on the level of funding for particular research fields within the organization, and on the training and recruitment rules for their own scientific staff. Weakly regulated institutional environments, and thus considerable organizational autonomy in national research systems, exist in the United States and the United Kingdom, whereas Germany and France serve as examples of strongly regulated environments. In the latter two countries, control over universities and public research organizations has been exercised to a large extent by state ministries. Therefore, decisions are typically made at state level, leaving little space for universities and research institutes to maneuver.

Hollingsworth’s (2004, 2006, p. 425–426) theoretical perspective has not yet been tested empirically with a larger sample of countries or with a larger sample of research fields or different measures of scientific breakthroughs. Returning to the abovementioned autonomy scoreboard (Pruvot & Estermann, 2017), both the United Kingdom and France receive scores in strong support of Hollingsworth’s claims, and Germany is only partially covered. Yet, we are far from



suggesting that the autonomy scoreboard data are perfect; recent studies illustrate serious cross-national measurement problems (Aksnes, Sivertsen, et al., 2017; Sandstrom & Van den Besselaar, 2018). Therefore, scholars of quantitative science studies should invest time and resources in developing large-scale, longitudinal data sets with variables anchored in middle-range theories such as Hollingsworth's. Successful examples show that such efforts can bear fruit. James March's (1991) concept of "exploration versus exploitation" has ignited a whole stream of mostly quantitative-empirical studies that have produced cumulative social scientific knowledge (for an overview, see Gibson & Birkinshaw, 2004; Raisch & Birkinshaw, 2008).

#### AUTHOR CONTRIBUTIONS

Conceptualization: AJ, TH; Investigation and Case Comparison: AJ; Writing: AJ, TH.

#### COMPETING INTERESTS

The authors have no competing interests.

#### FUNDING INFORMATION

01PY13013A; Federal Ministry of Education and Research BMBF Germany. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

#### DATA AVAILABILITY

Not applicable.

#### REFERENCES

- Aagaard, K., Bloch, C., & Schneider, J. W. (2015). Impacts of performance-based research funding systems: The case of the Norwegian Publication Indicator. *Research Evaluation*, 24(2), 106–117.
- Abbott, A. (1988). *The system of professions. An essay on the division of expert labor*. Chicago: University of Chicago Press.
- Abbott, A. (1991). The future of professions: Occupation and expertise in the age of organisation. *Research in the Sociology of Organisations*, 8, 17–42.
- Abramo, G., & D'Angelo, C. A. (2015). The VQR, Italy's second national research assessment: Methodological failures and ranking distortions. *Journal of the Association for Information Science and Technology*, 66(11), 2202–2214.
- Abramo, G., & D'Angelo, C. A. (2016). Refrain from adopting the combination of citation and journal metrics to grade publications, as used in the Italian national research assessment exercise (VQR 2011–2014). *Scientometrics*, 109, 2053–2065.
- Abramo, G., D'Angelo, C. A., & Caprasecca, A. (2009). Allocative efficiency in public research funding: Can bibliometrics help? *Research Policy*, 38, 206–215.
- Adler, R., Ewing, J., & Taylor, P. (2009). Citation statistics: A report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS). *Statistical Science*, 24(1), 1–14.
- Aksnes, D. W., Sivertsen, G., van Leeuwen, T. N., & Wendt, K. K. (2017). Measuring the productivity of national R&D systems: Challenges in cross-national comparisons of R&D input and publication output indicators. *Science and Public Policy*, 44(2), 246–258.
- Ancaiani, A., Anfossi, A. F., Barbara, A., Benedetto, S., Blasi, B., ... Sileoni, S. (2015). Evaluating scientific research in Italy: The 2004–10 research evaluation exercise. *Research Evaluation*, 24(3), 242–255.
- Anfossi, A., Ciolfi, A., Costa, F., Parisi, G., & Benedetto, S. (2016). Large-scale assessment of research outputs through a weighted combination of bibliometric indicators. *Scientometrics*, 107, 671–683.
- ANVUR. (2013). *Valutazione della qualità della ricerca 2004–2010 (VQR 2004–2010)*. Rapporto finale ANVUR Parte Prima: Statistiche e risultati di compendio. In Agenzia Nazionale di Valutazione del sistema Universitario e della Ricerca ANVUR.
- ANVUR. (2017). *Valutazione della qualità della ricerca 2011–2014 (VQR 2011–2014)*. Rapporto finale ANVUR Parte Prima: Statistiche e risultati di compendio. In Agenzia Nazionale di Valutazione del sistema Universitario e della Ricerca ANVUR.
- Bonaccorsi, A. (2018). *The evaluation of research in social sciences and humanities. Lessons from the Italian experience*. Heidelberg: Springer.
- Bonaccorsi, A., Cicero, T., Haddawy, P., & Hassan, S.-U. (2017). Explaining the transatlantic gap in research excellence. *Scientometrics*, 110, 217–241.
- Bornmann, L., & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, 66(11), 2215–2222.
- Cagan, R. (2013). The San Francisco Declaration on Research Assessment. *Disease Models & Mechanisms*, 6, Editorial.
- Capano, G. (2018). Policy design spaces in reforming governance in higher education: the dynamics in Italy and the Netherlands. *Higher Education*, 75, 675–394.

- Cimini, G., Zaccaria, A., & Gabrielli, A. (2016). Investigating the interplay between fundamentals of national research systems: Performance, investments and international collaborations. *Journal of Informetrics*, 10(1), 200–211.
- Espeland, W. N., & Stevens, M. L. (1998). Commensuration as a social process. *Annual Review of Sociology*, 24, 313–343.
- Geuna, A., & Piolatto, M. (2016). Research assessment in the UK and Italy: Costly and difficult, but probably worth it (at least for a while). *Research Policy*, 45, 260–271.
- Gibson, C. B., & Birkinshaw, J. (2004). The antecedents, consequences, and mediating role of organizational ambidexterity. *Academy of Management Journal*, 47(2), 209–226.
- Hicks, D. (2012). Performance-based university research funding systems. *Research Policy*, 41, 251–261.
- Hicks, D., Wouters, P., Waltman, L., de Rijke, S., & Rafols, I. (2015). The Leiden manifesto for research metrics. *Nature*, 520, 429–431.
- Hollingsworth, J. R. (2004). Institutionalizing excellence in biomedical research: The case of the Rockefeller University. In D. H. Stapleton (Ed.), *Creating a tradition of biomedical research. Contributions to the history of the Rockefeller University* (pp. 17–63). New York, NY: Rockefeller University Press.
- Hollingsworth, J. R. (2006). A path-dependent perspective on institutional and organizational factors shaping major scientific discoveries. In J. T. Hage & M. Meeus (Eds.), *Innovation, science, and institutional change* (pp. 423–442). Oxford: Oxford University Press.
- Jappe, A. (2019). Professional standards in bibliometric research evaluation? Results from a content analysis of evaluation studies in Europe. In *Proceedings of the 17th International Conference on Scientometrics & Informetrics ISSI, September 2–5, 2019* (pp. 1612–1623). Rome.
- Jappe, A., Pithan, D., & Heinze, T. (2018). Does bibliometric research confer legitimacy to research assessment practice? A sociological study of reputational control, 1972–2016. *PLOS ONE*, 13(6), e0199031.
- Kulczykcki, E. (2017). Assessing publications through a bibliometric indicator: The case of comprehensive evaluation of scientific units in Poland. *Research Evaluation*, 26(1), 41–52.
- Leydesdorff, L., Wagner, C. S., & Bornmann, L. (2014). The European Union, China, and the United States in the top-1% and top-10% layers of most-frequently cited publications: Competition and collaborations. *Journal of Informetrics*, 8, 606–617.
- March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, 2(1), 71–87.
- Merton, R. K. (1968). On sociological theories of the middle range. In R. K. Merton (Ed.), *Social theory and social structure* (pp. 39–72). Glencoe: Free Press.
- Miller, P. (2001). Governing by numbers: Why calculative practices matter. *Social Research*, 68(2), 379–396.
- Moed, H. (2017). *Applied evaluative informetrics*. Heidelberg: Springer.
- Molas-Gallart, J. (2012). Research governance and the role of evaluation: A comparative study. *American Journal of Evaluation*, 33(4), 583–598.
- NSB. (2019). *Science & engineering indicators–2018*. Arlington, VA: National Science Board.
- Petersohn, S. (2016). Professional competencies and jurisdictional claims in evaluative bibliometrics: The educational mandate of academic librarians. *Education for Information*, 32(2), 165–193.
- Petersohn, S., & Heinze, T. (2018). Professionalization of bibliometric research assessment. Insights from the history of the Leiden Centre for Science and Technology Studies (CWTS). *Science and Public Policy*, 45, 565–578.
- Porter, T. M. (1995). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.
- Power, M. (1997). *The audit society: Rituals of verification*. Oxford: Oxford University Press.
- Pruvot, E. B., & Estermann, T. (2017). *University autonomy in Europe III. The scorecard 2017*. Brussels: European University Association.
- Raisch, S., & Birkinshaw, J. (2008). Organizational ambidexterity: Antecedents, outcomes, and moderators. *Journal of Management*, 34(3), 375–409.
- Rottenburg, R., Merry, S. E., Park, S.-J., & Mugler, J. (2015). *The world of indicators. The making of governmental knowledge through quantification*. Cambridge: Cambridge University Press.
- Sandstrom, U., & Van den Besselaar, P. (2018). Funding, evaluation, and the performance of national research systems. *Journal of Informetrics*, 12(1), 365–384.
- Todeschini, R., & Baccini, A. (2016). *Handbook of bibliometric indicators: Quantitative tools for studying and evaluating research*. Weinheim: Wiley.
- Van Der Meulen, B. J. R. (2010). The Netherlands. In D. Simon, A. Knie, S. Hornbostel & K. Zimmermann (Eds.), *Handbuch Wissenschaftspolitik* (pp. 514–528). Wiesbaden: VS Verlag für Sozialwissenschaften.
- van Drooge, L., Jong, S., Faber, M., & Westerheijden, D. F. (2013). Twenty years of research evaluation. In *Facts & Figures*. The Hague: Rathenau Institute.
- van Eck, N. J., Waltman, L., Van Raan, A. F. J., Klautz, R. J. M., & Peul, W. C. (2013). Citation analysis may severely underestimate the impact of clinical research as compared to basic research. *PLOS ONE*, 8(4), e62395.
- van Steen, J., & Eijffinger, M. (1998). Evaluation practices of scientific research in the Netherlands. *Research Evaluation*, 7(2), 113–122.
- VSNU, KNAW, & NWO. (2003). *Standard Evaluation Protocol 2003–2009. Protocol for research assessments in the Netherlands*. Utrecht: Association of Universities in the Netherlands, the Netherlands Organisation for Scientific Research, and the Royal Netherlands Academy of Arts and Sciences.
- VSNU, KNAW, & NWO. (2009). *Standard Evaluation Protocol 2009–2015. Protocol for Research assessments in the Netherlands*. Available at: [www.knaw.nl/sep](http://www.knaw.nl/sep). Association of Universities in the Netherlands, the Netherlands Organisation for Scientific Research, and the Royal Netherlands Academy of Arts and Sciences.
- VSNU, KNAW, & NWO. (2014). *Standard Evaluation Protocol 2015–2021. Protocol for research assessments in the Netherlands*. Association of Universities in the Netherlands, the Netherlands Organisation for Scientific Research, and the Royal Netherlands Academy of Arts and Sciences.
- Whitley, R. (2000). *The intellectual and social organization of the sciences*, 2nd ed. Oxford: Oxford University Press.
- Whitley, R. (2007). Changing governance of the public sciences. In R. Whitley & J. Gläser (Eds.), *The changing governance of the sciences* (pp. 3–27). Dordrecht: Springer.
- Wildson, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., ... Johnson, B. (2015). *The metric tide: Report of the independent review of the role of metrics in research assessment and management*. Bristol: Higher Education Funding Council for England.