LUND UNIVERSITY

The heterogeneous landscape of bibliometric indicators: Evaluating models for allocating resources at Swedish universities

Hammarfelt, Björn; Nelhans, Gustaf; Eklund, Pieta; Åström, Fredrik

Published in:
Research Evaluation

DOI:
10.1093/reseval/rvv040

2016

Link to publication

Citation for published version (APA):
Hammarfelt, B., Nelhans, G., Eklund, P., & Åström, F. (2016). The heterogeneous landscape of bibliometric indicators: Evaluating models for allocating resources at Swedish universities. *Research Evaluation*, *25*(3), 292-305. https://doi.org/10.1093/reseval/rvv040

Total number of authors:
4

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Research Evaluation Advance Access published January 11, 2016

Research Evaluation, 2016, 1–14
doi: 10.1093/reseval/rvv040
Article

OXFORD

The heterogeneous landscape of bibliometric indicators: Evaluating models for allocating resources at Swedish universities

Björn Hammarfelt[1,2,*], Gustaf Nelhans[1], Pieta Eklund[1] and Fredrik Åström[3]

[1]Swedish School of Library and Information Science (SSLIS), University of Borås, 501 90 Borås, Sweden, [2]CWTS, Leiden University, 2333 AL Leiden, The Netherlands and [3]Lund University Library, P.O. Box 3, SE-221 00 Lund, Sweden

*Corresponding author. Email: bjorn.hammarfelt@hb.se.

Abstract

The use of bibliometric indicators on individual and national levels has gathered considerable interest in recent years, but the application of bibliometric models for allocating resources at the institutional level has so far gathered less attention. This article studies the implementation of bibliometric measures for allocating resources at Swedish universities. Several models and indicators based on publications, citations, and research grants are identified. The design of performance-based resource allocation across major universities is then analysed using a framework from the field of evaluation studies. The practical implementation, the incentives as well as the 'ethics' of models and indicators, are scrutinized in order to provide a theoretically informed assessment of evaluation systems. It is evident that the requirements, goals, possible consequences, and the costs of evaluation are scarcely discussed before these systems are implemented. We find that allocation models are implemented in response to a general trend of assessment across all types of activities and organizations, but the actual design of evaluation systems is dependent on size, orientation, and the overall organization of the institution in question.

Key words: research evaluation; performance indicators; university governance; bibliometrics; Sweden.

1. Introduction

Bibliometric measurement is an integrated part of everyday practices at many universities. Each university is ranked every year in various rankings, the output of specific departments is studied and compared in internal evaluations, and individual researchers are repeatedly assessed through bibliometric measures such as the h-index or the Journal Impact Factor. In this study, we focus on one specific use of bibliometric indicators: the systematic measurement of publications and citations for resource allocation *within* universities.

It could well be that bibliometric measures applied to a university, or in some instances even at faculty or departmental level, might have more impact on the research practices of the individual scholar than the national systems for allocating resources across universities. In this sense, we agree with Burrows (2012: 359)in his view that all bibliometric measures '…are nested or folded into each other to form a complex data assemblage that confronts the individual academic'. Thus, research on the use and effects of evaluation using such measures should include a focus not only on the national (macro level) or individual (micro level) but also on the institutional (meso level). Yet, the use of bibliometric measures—for allocation, evaluation, and promotion—within universities has rarely been studied, and few structured overviews exist.

A consistent implementation of bibliometric measures all the way down through the implementation chain is often presumed when studying national evaluation and resource allocation systems. However, even in national contexts dominated by one specific indicator, it has been found that the implementation of bibliometric measures varies significantly across institutions (Aagaard 2015).

© The Author 2016. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Knowledge about the local and institutional use of indicators then emerges as crucial to our overall understanding of the ways in which evaluation systems and resource allocation systems influence research output and practices. Consequently, the main purpose of our study is to investigate how national incentives and models trickle down and translate on a local university level.

First, however, we must map and describe the various models and indicators currently applied locally at Swedish universities. All Swedish higher education institutions (HEIs) awarding third-cycle degrees (PhDs) were included in the study. In total, we targeted 27 of the major HEIs in Sweden, and 26 responded. Using this sample, we provide a first, and up-to-date, mapping of the heterogeneous, and ever-changing, landscape of resource allocation models across Swedish universities.

In the second part of the study, we systematically evaluated the allocation models used by universities in our sample. Bibliometric indicators are often scrutinized, criticized, and revised, but a broader evaluation, which goes beyond methodological refinement, is seldom applied. Given the current focus on auditing and evaluation, which is illustrated by notions of an audit (Power 1997) or evaluation society (Dahler-Larsen 2012a), it is surprising that evaluation systems and indicators seldom are the subjects of evaluation themselves. By utilizing a framework that is rooted in the field of evaluation studies, we provide a perspective that is rarely applied in research on bibliometric indicators. Research on technical and methodological aspects is indeed of great relevance, but such a narrow focus has to be complemented in order to provide a thicker description of bibliometric measurements and their role in academia. Therefore, we used and adapted a theoretical frame drawn from the work of political scientist, Peter Dahler-Larsen (2012a,b), designed to 'evaluate evaluation'. His theory is not specifically developed to analyse bibliometric measures, but it offers a basis for understanding bibliometric performance indicators as part of a general trend of 'evaluation machines' rather than focusing on the specific methods or data sources used.

With the implementation of national systems using bibliometric measures, it can be expected that institutions and individuals will respond strategically: 'The academics are very creative—they can and will respond to such measures in novel and unforeseen ways. And the institutions they serve will always seek to maximise returns, as is their responsibility'. (Butler 2010: 158). The local evaluation systems reviewed in this study could in many ways be seen as strategic responses in order to 'maximise returns'. The current study should also be viewed in the light of recurring calls for guidelines and standards for bibliometric assessment that have recently been voiced within the bibliometric community (Gingras 2014; Hicks et al. 2015). Moreover, this study is also related to an on-going debate in Sweden regarding the use of bibliometric measures on the national level (Nelhans 2013). Similarly, the interest in how bibliometric evaluation affects research practices is growing (Woelert and Yates 2015; De Rijcke et al. in press), and, although this study does not concern itself with effects, hopefully it provides an overview of the systematic use of indicators, which can help us understand their impact on strategies and practices. Thus, we deliberately choose to evaluate these models based on how they are designed and formally described rather than on their actual implementation, use, and effects. In doing so, we make use of specific evaluation criteria, formulated by Dahler-Larsen (2012b), which can be summarized under three different headings: *Legitimacy and appropriateness*, *Organizational and methodological stability*, and *transparency, feedback, and learning*. Such an approach allows us to go beyond the critique of specific indictors and formulate a broader framework for evaluating these systems. Finally, the article engages with the question of the institutionalization of bibliometric measurement and, in accordance with this perspective, we adhere to the view of publication and citation databases as well as systems of evaluation as part of a growing evaluative 'infrastructure' (Wouters 2014).

The article is organized as follows. First, a concise description of the use of performance-based resource allocation in Sweden is provided. Then, we briefly review current literature that critically engages with the implementation of bibliometric evaluation. After that, we outline our theoretical framework, and the three main criteria for assessing bibliometric evaluation systems are presented. Subsequently, we introduce and describe our sample and the methods used for collecting our data. Next, we present our findings both with the intention of overviewing indicators used generally as well as a more detailed and systematic evaluation of specific models. Our discussion elaborates on the empirical findings and deliberates on the potential that an 'evaluative perspective' entails for bibliometric research.

2. Background

Performance-based research funding systems (PRFS) have been introduced in numerous countries since the mid-80s and onwards. Australia was the first nation to impose a general and comprehensive system for resource allocation based on bibliometric measures. Since then, we have seen a range of different systems using either, or both, publication and citation counts. While national models have attracted the attention of researchers interested in their construction (Schneider 2009; Ahlgren et al. 2012), or in the debate that arose on their introduction (Nelhans 2013), there has been little research on the use of bibliometric measures at the institutional level.

PRFS can be defined using four criteria adapted from Hicks (2012): (1) research must be evaluated (thus, teaching and other activities are excluded), (2) research evaluation must be done *ex post* (not *ex ante* as in the case of program or project funding), (3) research output must be evaluated (systems focusing only on numbers of PhD students or incoming grants are excluded), and (4) government distribution of research funding must depend on the results of evaluation. Hicks (2012: 252)also included a fifth criterion stating that PRFS must be applied on the national level. However, we see no reason to limit the definition only to the inclusion of assessment on the national level, as the concept and basic construction of PRFS remains the same even on the institutional level.

In her overview, Hicks (2012) found that 14 countries had launched or were planning to launch PRFS in 2010. The rationale for PRFS is simple: institutions that perform according to the quality criteria formulated in the system should be awarded a larger share of the available resources. Two main reasons for implementing such a system can be found in the literature on the topic. One argument, often evoked by proponents of innovation and the knowledge economy, focuses on globalization and competitiveness. Researchers interested in developments within higher education take a contrasting position and PRFS are viewed in the light of accreditation, assessment, and new public management (Hicks 2012: 253). Our study is firmly situated in the latter category, as we study the institutional response to the growing trend of systematically evaluating publicly funded activities.

2.1 The Swedish system and rivalling models for allocation of resources across universities

PRFS is a quite recent phenomenon in Swedish universities. The introduction of a new allocation system on the national level, as well as the administrative trend of using 'output measures' in public management, provided incentives for developing local models for resource allocation.

The current Swedish system for the allocation of resources uses two indicators: a bibliometric indicator allocating resources based on the number of publications and citations, and an indicator based on the amount of external funding acquired by each university (Prop. 2008/09:50 2008). Ten per cent of the annually awarded funds, which increased in 2014 to 20%, were allocated to Swedish HEIs based on these indicators in equal parts (Fig. 1).

The system was deliberately constructed to provide 'strong incitements to increase activity on the global publication market' (SOU 2007:81:418), and to encourage high-quality research rather than the production of numerous low-quality publications, although for the humanities and the 'soft' social sciences, the goal was to increase overall output (Sandström and Sandström 2009: 249). The model immediately met with critique. First, the model in itself was not robust enough; and second, the social sciences and the humanities performed poorly as a result of being evaluated by measures not suited for their respective publication and citation patterns. In the end, the Swedish government decided to use the proposed model, albeit with two modifications. For the humanities, the citation factor was set to 1 instead of the actual figures entailing that citations did not count at all for this area. Furthermore, the government arbitrarily added an 'extra weighting factor' multiplying the indicator values for different research areas according to a set of variables that had gained long-standing acceptance in Swedish research policy: *Medicine and technology* 1.0, *(natural) science* 1.5, *social sciences and humanities* 2.0, and *other areas* 1.1 (Prop. 2008/09:50: 57).

The model was used the first time in the allocation of funds for Swedish HEIs in 2009, and the effects of the Swedish model on research practices and strategies are yet to be revealed, although initial findings suggest that at least some modest changes have occurred (Hammarfelt and de Rijcke 2015). However, the influence of any single system cannot be assessed without an understanding of the broader context in which it is employed. We therefore agree with Hicks' (2012: 259) assertion that '… the influence of a [national] PRFS will depend on how universities allocate funding internally; conceivably university management could negate or enhance PRFS incentives'.

It should also be mentioned that in 2013, the Swedish Government commissioned the Swedish Research Council to investigate an allocation system based on peer-review panels instead of bibliometric indicators, where the performance-based allocation of funds would be based on scientific (or artistic) quality (70%), quality enhancing factors (15%), and impact outside academia (15%). However, bibliometric measures are not discouraged totally in the model, as data on citations will inform panels in research areas where it is deemed 'appropriate'. Should the Government accept the proposed model, implementation is planned for 2017–18, and a first round of fund allocations would take place in 2019 (Forskningskvalitetsutvärdering i Sverige 2014).

The Swedish model only assesses and distributes resources across universities, and in this sense, it differs from other models, such as the the British Research Excellence Framework (REF), which evaluate and publish reports on the performance of individual departments while reallocating resources on the university level. Others, such as the Spanish Sexenio and the New Zealand Performance Based Resource Fund (PBRF), evaluate and distribute resources on the level of individuals (Hicks 2012: 254). The designers of the Swedish model explicitly state that the model is only supposed to be applied on a national level, and use *within* universities is discouraged. Yet, as will be evident from this study, incentives and models applied on the macro level have a tendency to trickle down and confront the individual researcher.

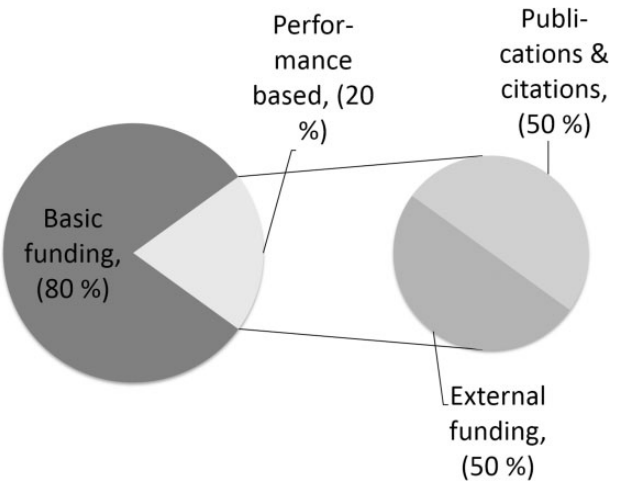

Figure 1. Swedish model for resource allocation.

2.1.1 The Norwegian model

Versions of the Norwegian model are currently used at several Swedish universities, warranting a short introduction to this system. The Norwegian model was developed for performance-based funding across universities, not within them (Sivertsen 2008). The system is designed to capture the total output of publications; monographs, chapters in edited books as well as journal articles are all viable publication channels as long as the work has been peer reviewed. All publication channels are rated using three levels: 'unscientific', 'scientific', and 'prestigious', corresponding to Levels 0, 1, and 2, respectively. Only publications at Level 1 (amounting to roughly 80% of items published in ranked sources in each discipline) and Level 2 (~20 of the items) are awarded in the redistribution of funds. The categorization of publications as Level 1 or 2 is carried out in research councils, but for eligible fields (foremost natural sciences and medicine), the lists are partly based on journal impact factors derived from Web of Science (Schneider 2009). Publications are then awarded points based on publication channel and level (Table 1).

An obvious weakness in the Norwegian system is that a two-level rating is an unsophisticated measure of the quality of a publication. However, in the context of the model being applied locally at Swedish universities, we find that a more important critique is the fact that it is Norwegian. The inclusion and rating of publications is specifically developed for conditions in Norway, and important channels in a Swedish context might not be categorized at all, or at a level that does not reflect its standing within the academic community.

This brief overview has focused on bibliometric measures; yet, peer-review procedures as well as other measures of funding, graduates, internationalization, and innovation play an important role in many evaluation systems. Non-bibliometric measures also play a significant role in many of the institutional systems studied here, and the acquisition of external grants is one indicator that is often

Table 1. The division of channel and level in the Norwegian model based on publication channel and quality level

Publication channel	Level 1	Level 2
Monograph (ISBN)	5	8
Article in periodic outlet or series (ISSN)	1	3
Article in anthology	0.7	1

employed. It should therefore be noted that the focus of this study is solely on the use of bibliometric indicators, and the specificities of other types of indicators are left uncharted.

2.2 The rise of a metric culture in academia

The increased use of bibliometric measurements across academia, described as a 'profusion of measures' (Van Noorden 2010), 'metric assemblages' (Burrows 2012), or the rise of a 'metric culture' (Hammarfelt and de Rijcke 2015), has led to greater focus on developing guidelines for their use. Recent examples of this growing trend are the guidelines outlined by Gingras (2014) for developing indicators, and we also see an emerging interest in the ethics of bibliometric evaluation (cf. Furner 2014).

The consequences of an increasing use of bibliometric measurements for the evaluation of universities and researchers have been discussed for some time, and literature with a critical approach to bibliometric evaluation is steadily growing. A pioneering study of the effects of bibliometric measurement is Butlers' (2003) analysis of publication patterns in Australian universities. She was able to show that the implementation of a new publication-based PRFS in Australian higher education led to more publications but impact did not increase; rather, she revealed a tendency towards lower impact as measured by citation counts. Ossenblok et al. (2012) also found that researchers in Flanders and Norway reacted to the incentives by publishing more in top international journals of their respective evaluation systems. Similar results were seen in Denmark, where the introduction of a national PRFS appeared to have positive effects on the production of research articles (Ingwersen and Larsen, 2014). Later studies have complemented these findings with more theoretically informed discussions of how bibliometric measures may influence academic research on a broad scale (Strathern 2000, Weingart 2005; Burrows 2012; Lorenz 2012). Studies, such as those of Woelert and Yates (2015), add a different perspective, where evaluation systems renegotiate values pertaining to trust and authority within the university system. The effects that formalized ex post evaluation might have across disciplines have also been theoretically described by Whitley (2007).

One of few accounts of bibliometric evaluation within a specific university is given by Hammarfelt and de Rijcke (2015) in their study of the allocation system used at the faculty of Arts at Uppsala University (UU). This model is complex, as it distributes resources based on four indicators: external grants, grants from the Swedish Research Council, citations from Web of Science, as well as points based on the output of publications (using the Norwegian list). Some diverging views could be heard in the discussion prior to the introduction. The decision to implement this system was taken in 2011, and the introduction of a new system for allocation of resources provided an incentive for a local model. Systematic changes in publication practice in response to the model's incentives could be seen; among the more notable were an increase in English language publication as well as a considerable increase in the number of peer-reviewed publications. Researchers also explicitly commented on how bibliometric measures and demands from funders influence research practices.

Common to many of these studies is the effort to distinguish between the 'unintended' or 'inadvertent' effects of bibliometric evaluation. Yet, we concur with Dahler-Larsen (2012a) that a focus on unintended effects may reduce and distort our understanding of how these measures influence knowledge production. An important observation is that in order to discern unintended effects, we must first be able to formulate the intended effect of a specific system. In many contexts, the intended effects of evaluation are not clearly distinguishable or are so vaguely formulated that it is advisable to focus on constitutive rather than unintended effects (Dahler-Larsen 2012a: 201–2). Such an approach is less restrictive, and it opens up for an analysis of the 'performative functions' of bibliometric measurement (Burrows 2012; Nelhans 2013).

The studies and frameworks discussed above mainly focus on the practical and methodological aspects of applying bibliometric methods rather than on the act of evaluation as such. Thus, although these guidelines are much needed, we propose that the whole evaluative process—including the choice to evaluate in the first place—should be part of any critical study of assessment regimes. In doing so, we suggest that frameworks from the field of evaluation studies could further our understanding of how bibliometric assessment is used. Such a framework, developed by Dahler-Larsen (2012b), is introduced in the next section.

3. A theoretical framework for assessing research evaluation systems

A fundamental premise for this study is that performance-based resource allocation can be defined as a type of evaluation. As with all complex concepts, there are several partly overlapping definitions of evaluation, and these are related to the approach taken. Dahler-Larsen (2012a) identifies three main perspectives—conceptual-analytical, methods-focused, and purpose-focused—all of which incorporate slightly different definitions and operationalizations of the concept. However, most definitions include four aspects of evaluation: '(1) an evaluand, (2) some assessment based on some criteria, (3) a systematic approach or methodology to collect information about how the evaluand performs on these criteria, and (4) a purpose or intended use'. (Dahler-Larsen, 2012a: 9). In line with these aspects, he provides a definition of evaluation: '[. . .] evaluation is basically a systematic, methodological, and thus "assisted" way of investigating and assessing an activity of public interest in order to affect decisions or actions concerning this activity or similar activities'. (Dahler-Larsen 2012a: 9). Thus, we argue that performance-based allocation models can and should be defined as a type of evaluation. They include an evaluand (research, researchers); criteria are established (publications, citations, and grants received); they are systematically used; and they have a specific purpose (increase impact/quality of research). It could be maintained that evaluation in this context is more of a measurement than a judgment, but the reduction of these systems as merely 'neutral' administrative systems designed to allocate rather than judge disguises their function as devices that designate value and comparability. However, it should also be noted that the allocation models studied here encompass a particular type of evaluative activity that is specifically designed to provide incentives through the allocation of resources. The type of resources allocated—research time, travel money, or other funds—is

therefore an additional component to consider when analysing these systems. Moreover, the size and proportion of resources allocated based on performance is another factor that should be considered in the analysis.

Performance-based resource allocation is an evaluative practice that is formalized and repeated. These characteristics allow us to define them as 'evaluation systems' as opposed to single event assessments. Thus, in order to qualify as an evaluation system, it must be permanent, routinized, and extended across time and space. Furthermore, evaluation systems, compared to 'regular' evaluation, are less '[. . .] dependent on the values and ideas and styles of individual evaluators'. (Dahler-Larsen 2012b: 31). Rather than being dependent on individual judgements—as is the case with peer-review procedures—evaluation systems represent institutional values, and they use tools such as indicators, criteria, and standards.

Based on our understanding of performance-based allocation as evaluation systems, we have outlined three main criteria for assessing models for bibliometric evaluation at the selected universities: (1) *Legitimacy and appropriateness of the system*, (2) *Organizational and methodological soundness and stability*, and (3) *Degree of transparency and learning*. These criteria are deduced from an outline of 19 questions for evaluating evaluation systems developed by Dahler-Larsen (2012b). Initially, we grouped these questions into three themes, which roughly correspond with the criteria above. However, by reducing and merging the 19 questions under three main headings, rather than presenting a long list of questions of which several could be answered by a simple yes and no, we aim to provide a framework that captures the wide range of criteria needed to assess evaluation system while at the same time offering a clear and distinctive analysis of PRFS currently in use at Swedish universities. Furthermore, Dahler-Larsen's framework was initially constructed to assess evaluation *a priori*. Thus, a modification and reworking of the framework was required in order to use it for evaluating systems that are already in use.

3.1 Legitimacy and appropriateness

This criterion is used to examine the rationale for implementing an evaluation system and to consider if the method of evaluation (e.g. the indicators used) corresponds with the activity it is supposed to assess. It is also used to discern whether the evaluated activity is important enough to motivate the introduction of an evaluation system. Other crucial questions are if the activities being evaluated are adequately represented by the indicators used and if the goals of the activity are agreed upon. This criterion also engages in the issue of micro accountability and scrutinizes the principal ideology behind the system. Finally, it poses the question of how the evaluated are likely to behave if they take the incentives in the model seriously. Thus, if the criteria in the model are met, is the behaviour good, and vice versa? (Dahler-Larsen 2012b). Generally this criterion targets the rationale and motivation for implementing research evaluation systems, and many of the issues raised are concerned with the decision to implement a specific system.

3.2 Organizational and methodological stability

The second criterion concerns the stability and reliability of the actual infrastructure supporting the evaluation systems. Technical and methodological issues are in focus, as well as the system's capacity for providing reliable and trustworthy information. How the evaluation system is anchored in the organizational structure and how mandatory it is for all fields is another question of great interest (Dahler-Larsen 2012b). These questions target the infrastructure (databases and administrative systems) used in evaluations, and it also concerns the personnel conducting bibliometric analyses.

3.3 Transparency, feedback and learning

The final criterion concerns the awareness of scholars about the system and feedback from it. Learning mechanisms and responsiveness are here seen as important aspects of an evaluation system. The reflexivity of the system also concerns its own construction and implementation. Has the system been tested before implementation or assessed in practice? And are the costs of implementing and undertaking evaluation well described? (Dahler-Larsen 2012b). How information about these systems is communicated might also have consequences for the effects that they have. If scholars have little knowledge about the system's incentives, then we could perhaps assume that these measures have little effect on publication patterns and research practices.

4. Methodology

There are 47 HEIs in Sweden according to the Swedish Higher Education Authority. As our focus is on research, we limited our sample to HEIs that can award third-cycle degrees (PhDs), as little research is conducted at other institutions. Our sample of 27 HEIs includes all state-controlled large and small universities as well as independent HEI entitled to award first-, second-, and third-cycle qualifications (Table 2). This selection includes a variety of institutions, including large multidisciplinary universities such as UU and University of Gothenburg (GU) to smaller regional universities such as Halmstad University (HH) and the University of Borås (HB) as well as specialized institutions such as Chalmers University of Technology (CTH).

4.1 Questionnaire

The selected institutions were contacted through personnel responsible for issues concerning bibliometrics and research evaluation, and were asked to respond to eight questions through e-mail (see Appendix A). A first round of questionnaires was sent out in spring, 2014, collecting responses from 13 HEIs. In the autumn, 2014, this initial sample was extended to include all 27 universities. All HEIs responded to our questionnaire, with the exception of University West. Data collection for this project was carried out during the period May to September, and it is important to note that several HEIs had recently implemented new models, while others were in the process of developing or revising their models. The landscape of resource allocation models emerged as dynamic, and our map may very well be outdated in a few years' time. However, rather than being discouraged, we find that the dynamic development of indicators and systems is a further motivation for studying their construction and implementation.

4.2 Document search and analysis

Responses to our questionnaire, which range from thoughtful lengthy descriptions to short yes' and no's, were supplemented by official and semi-official documents regarding research allocation models. Some of these documents were kindly provided by the respondents themselves (see Question 6 in Appendix A), while others have been gathered through systematic searches of institutional websites. The documents have been examined in order to clarify the design of specific systems and the construction of impact indicators. Apart from being a source

Table 2. Universities included in the sample ranked according to number of doctorates conferred in 2012

HEI	Number of doctorates conferred 2012[a]	Profile	Steering
Lund University (LU)	325	General	Governmental
Uppsala University (UU)	314	General	Governmental
Karolinska Institutet (KI)	301	Medicine	Governmental
Royal Institute of Technology (KTH)	235	Techn. sciences	Governmental
University of Gothenburg (GU)	233	General	Governmental
Stockholm University (SU)	229	General	Governmental
Umeå University (UmU)	174	General	Governmental
Chalmers University of Technology (CTH)	172	Techn. sciences	Independent
Linköping University (LiU)	169	General	Governmental
Swedish University of Agricultural Sciences (SLU)	103	Agr. Sciences	Governmental
Luleå University of Technology (LTU)	57	General	Governmental
Örebro University (ÖU)	55	General	Governmental
Linneaus University (LNU)	38	General	Governmental
Karlstad University (KAU)	25	General	Governmental
Jönköping University (HJ)[b]	21	General	Independent
Mälardalen University (MDH)[b]	20	General	Governmental
Blekinge Institute of Technology (BTH)[b]	18	Techn. sciences	Governmental
Stockholm School of Economics (HHS)	16	Economics	Independent
Mid Sweden University (MIU)	14	General	Governmental
Malmö University (MAH)[b]	13	General	Governmental
University of Borås (HB)[b]	2	General	Governmental
Halmstad University (HH)[b]	1	General	Governmental
Södertörn University (SH)[b]	0	General	Governmental
The Swedish School of Sport and Health Sciences (GIH)[b]	0	Sport and health	Governmental
University of Gävle (HIG)[b]	0	General	Governmental
University of Skövde (HS)[b]	0	General	Governmental
University West (HV)[b]	0	General	Governmental

[a]Data from http://www.uka.se/download/18.1c251de913ecebc40e780003405/1403093616367/annual-report-2013-ny.pdf (accessed 18 February 2015).
[b]University Colleges entitled to award third-cycle qualifications in one or several disciplinary domains.

for information on the design and construction of the systems, official documentation provides an important source for analyzing the motivations for implementing specific systems. All documents used in our analysis are listed in Appendix B.

5. Mapping the Swedish landscape of bibliometric indicators and systems

Bibliometric measurements for resource allocation are currently employed at almost all major HEIs: 24 of 26 use bibliometrics at one or several levels within their organization, albeit with great variation in terms of extent; in some places, redistribution is managed throughout the university, and in others, the use of metrics is limited to certain faculties or research areas. Only CTH and the Stockholm School of Economics refrain entirely from employing such measures, although discussions on the implementation of a system for bibliometrics-based resource allocation at CTH are ongoing. This can be compared to previous studies, where a report from 2008 revealed that 13 of 38 HEIs in Sweden used bibliometric measures for evaluative or analytical purposes (Carlsson and Hällgren 2008). A study published in early 2013 revealed that 16 of 34 faculties at major universities use bibliometric measures for allocating resources (Görnerup 2013). Several respondents in our study stated that models were either recently employed or under development, and a substantial increase in the use of bibliometric measurements at Swedish HEIs is evident. Bibliometric measurements are now no longer the exception, but the norm.

Our findings show that a variety of different systems, using a range of indicators, models, and measures, are employed at Swedish universities (Table 3). These models differ considerably in their sophistication and complexity, but they all share the same vague goal of 'increasing quality of research'.

Generally, smaller universities use publication-based systems ranging from models using raw publication counts to variants of the Norwegian model. A couple of specialized, high status, institutions use citation-based models, while the majority of larger universities use combinations of publication and citation-based measures. Sometimes, these are integrated in one model, but it is also common that different models are used depending on the research domain; for instance, medical faculties may use citations counts, while the social sciences and humanities opt for publication-based models.

5.1. Different types of bibliometric indicators

So far, we have found that a range of measures, of different levels of complexity, are used across universities in Swedish academia: from raw publication points to more advanced indicators using percentiles and field normalization. Many institutions rely on well-known and established measures, although some of them contested, while others have developed their own models designed specifically for their needs. As illustrated above, two main types of bibliometric indicators have been identified in our sample—those based on publication counts and those based on citations.

Table 3. Main bibliometric input for allocation of resources

Publication based (11)	Citation based (2)	Combination of C and P (11)
Blekinge Institute of Technology	Karolinska Institutet	Jönköping University
Halmstad University	KTH	Karlstad University
Linneaus University		Linköping University
Luleå University of Technology		Lund University[a]
Mid Sweden University		Malmö University
Mälardalen University		Swedish University of Agricultural Sciences
Stockholm University		The Swedish School of Sport and Health Sciences
Södertörn University		Umeå University
University of Borås		University of Gothenburg[a]
University of Gävle		Uppsala University
University of Skövde		Örebro University

[a]At selected faculties.

5.1.1 Publication counts

The most straightforward method for assessing research performance is to count publications. Ten HEIs in our sample use publications as the only bibliometric component in their allocation model. Yet, the type of publication counted, the extent to which all are given the same points, or if a graded scale is applied differs considerably. Some, like Södertörn University (SH), use revised models based on the Norwegian system for allocation, while others have developed a method of their own. The complexity of these systems ranges from quite unsophisticated systems to more developed models that differentiate between both publication type and the 'quality' of the publication channel. HB uses a simple model in which peer-reviewed articles and 'total research output' are calculated based on data from the local publication database. The calculation of 'total research output' is not described, and the meagre description of the system gives us a few clues about how the system actually works (Styrmodell vid Högskolan i Borås, 2011). Similar models, counting publications in peer-reviewed journals, for instance, are used at Mälardalen University and at the University of Gävle.

A more complicated system is used at Linneaus University (LNU). Here, publication points based on the Norwegian model are normalized using a benchmark of other departments and universities of a similar size. Researchers at LNU also get points for publications in journals indexed in Web of Science and which are not included in the Norwegian system, while conference proceedings are awarded the same number of points as book chapters. HH and the University of Skövde use systems based on the same principles as the Norwegian system, although they allocate points to a wider range of publications. The model at HH awards 3 points to a journal article, 1 to conference proceedings, 2 to chapters in books, and 8 to monographs. Furthermore, extra points can be added—a journal article on Level 2 in the Norwegian system gets an additional 2 points; a peer-reviewed conference proceeding gets an extra point; and if it is indexed in the Norwegian system, it receives an additional 2 points; the same applies to book chapters (Modell för fördelning av forskningsbidrag till forskningsmiljöer, 2011). Most universities using more-complex models also fractionalize publication counts, and this also applies to universities using variants of the Norwegian model. Less-complicated models based on all publications or journal articles tend to use whole number counting.

5.1.2 Number of citations

Few HEIs use citations counts as their main bibliometric indicator. Karolinska institutet (KI) and The Royal Institute of Technology (KTH) are the only institutions in our model that only count citations. KI has its own database with access to Web of Science, allowing them to carry out regular assessments. The model used by KI is quite complex, and it uses three main indicators: journal impact factor, field-normalized citations, and total number of citations (Användning av bibliometri [...], 2009). KTH also uses field-normalized citations, and both KI and KTH use whole number counts. At KI, this is a strategic decision, as they explicitly state that whole counts are used to encourage collaboration. Both KI and KTH are specialized institutions, and their disciplinary profiles lend themselves to citation-based evaluation. These more-advanced indicators also require resources in the form of skilled bibliometricians and direct access to citation databases. The use of more-advanced citation-based evaluation might therefore be limited to specialized and reasonably large institutions.

5.1.3 Publications and citations

Many of the larger and more diversified universities use both publication counts and citation counts for resource allocation. Some models cover all university activities, while others use different indicators depending on the faculty and department being evaluated. Lund University is an example of a HEI where bibliometric measures are used at three faculties but not for the university as a whole. The Social Science faculty counts whole publications and awards points according to type, the Natural Sciences use field normalized citations delivered from the Centre for Science and Technology Studies, while the School of Economics uses fractionalized publications counts in combination with the Norwegian list. Umeå University (UmU) has a similar approach, where each faculty has chosen their own model, although allocation also takes place on a central level. Similar mixed-models are used at UU and GU. Thus, large universities usually adapt mixed systems to encompass specific models attuned to their own disciplinary traditions. The strong position of the faculties in the administrative structure of these universities also allows for greater independence in the choice of evaluation system.

The Swedish Agricultural University (SLU) is an example of a university with a mixed allocation system for the whole organization. Here, a journal article is awarded one point if it is indexed in the local database SLUPub, it receives an extra half point if it is published in a journal with a higher impact factor than the average for the research field, and another half point if it is among the top 25% most cited in the field. Books receive 4 points, while book chapters are awarded 0.5 points. This is one of the more elaborated models we encountered,

Table 4. Swedish HEIs and the use of bibliometrics on various levels

HEI	Faculties (10)	Departments (16)	Individuals (6)
Blekinge Institute of Technology			X
Jönköping University	X (professional schools)	X	
Karlstad University		X	X
Karolinska Institutet		X	
KTH	X (schools)		
Linköping university		X (Health Science)	
Linneaus University	X		X
Lund University		X[a]	
Malmö University	X	X	
Mid Sweden University		X	
Mälardalen University	X (research spec)		
Luleå University of Technology		X	X
Stockholm University		X	
Swedish University of Agricultural Sciences	X	X (not formalized)	
Södertörn University		X	X
The Swedish School of Sport and Health Sciences			X
University of Borås	X		
University of Gothenburg		X[a]	
University of Gävle		X	
University of Halmstad		X (research area)	
University of Skövde		X	
Umeå University	X	X	X
Uppsala University	X[a]	X	
Örebro University	X		

[a]At selected faculties.

and it is also notable that it combines features of the Norwegian system with citation-based measures in a unique system of evaluation.

5.2 Level of application

Performance-based resource allocation takes place at many different levels at Swedish HEIs. It is used to redistribute resources across faculties, departments, and individuals (Table 4). Most universities use bibliometric measures for allocating resources to faculty or at departmental level; however, there are also quite a few universities that assess researchers on the individual level. We suspect that bibliometric measures trickle down to the level of the individual researcher in this way much more often than is indicated in our survey. It is also possible that administrators use and refer to such measures in discussions regarding employment and salary. Furthermore, our study has been conducted mainly at the level of faculties and departments and a part from a few hints we know little about how bibliometric indicators are used within specific departments. Even less is known about the informal use of bibliometric indicators in assessing, employing, and recruiting academic personnel. Thus, although the topic of this study is the organized use of bibliometric measures for allocating resources, bibliometric measurements may not be limited to its systematic and formalized use.

Typically, the bibliometric community of researchers and analysts declare that bibliometric measurements should be used on an aggregated level only, whereas its use in assessing individuals has been discouraged. Still, a number of measures have been developed for evaluating individual scholars (Wildgaard et al. 2014). The rather common use of bibliometric measures for the allocation of resources to individual researchers is therefore of particular interest. Blekinge Institute of Technology (BTH), Karlstad University, LNU, University, Luleå University of Technology (LTU), SH, The Swedish School of Sport and Health Sciences, and UMU use individual-level bibliometrics, either throughout the whole university or in one or more research areas, for allocating resources to individual scholars or scientists. Some of these systems are quite straightforward—authors are directly awarded in cash for articles published in peer-reviewed journals. A scholar at LTU gets 35,000 SEK (about 3,800 EURO) for a peer-reviewed journal article. Moreover, if the article is indexed in Web of Sciences or published in a Level 2 journal in the Norwegian system, the author is rewarded an additional 35,000 SEK. (Publiceringsstöd till vetenskapliga artiklar och konstnärliga produktioner, 2014). BTH has a similar system in which *Web of Science*-indexed articles are rewarded with 30,000 SEK. The system at LNU is slightly more intricate, as researchers are awarded publications points based on a revised version of the Norwegian system that later are translated into resources. However, if the publication points gathered by a researcher are worth less than 8,000 SEK, these points are awarded to the department instead. Furthermore, researchers in the top 20% of earners of points are given an additional 15,000 SEK. Hence, this system is deliberately designed to encourage highly productive researchers and to punish unproductive ones.

In order to illustrate how systems and indicators in different models interact with each other, we have chosen to zoom in on one particular example, UmU. UmU is selected, as it currently uses bibliometric evaluation on all three levels analysed in this study (Fig. 2).

The many levels and indicators involved illustrate the complexity of these systems. Interestingly, these indicators represent different kinds of incentives, which in some situations, might contradict each other. If we add other layers of bibliometric measurement, such as university rankings and h-index of individual scholars, the picture becomes even more complicated. Thus, the example of UmU reflects that universities are heterarchical organizations when it comes to evaluation; no single evaluation criterion or indicator can capture the breadth of research activities, and consequently,

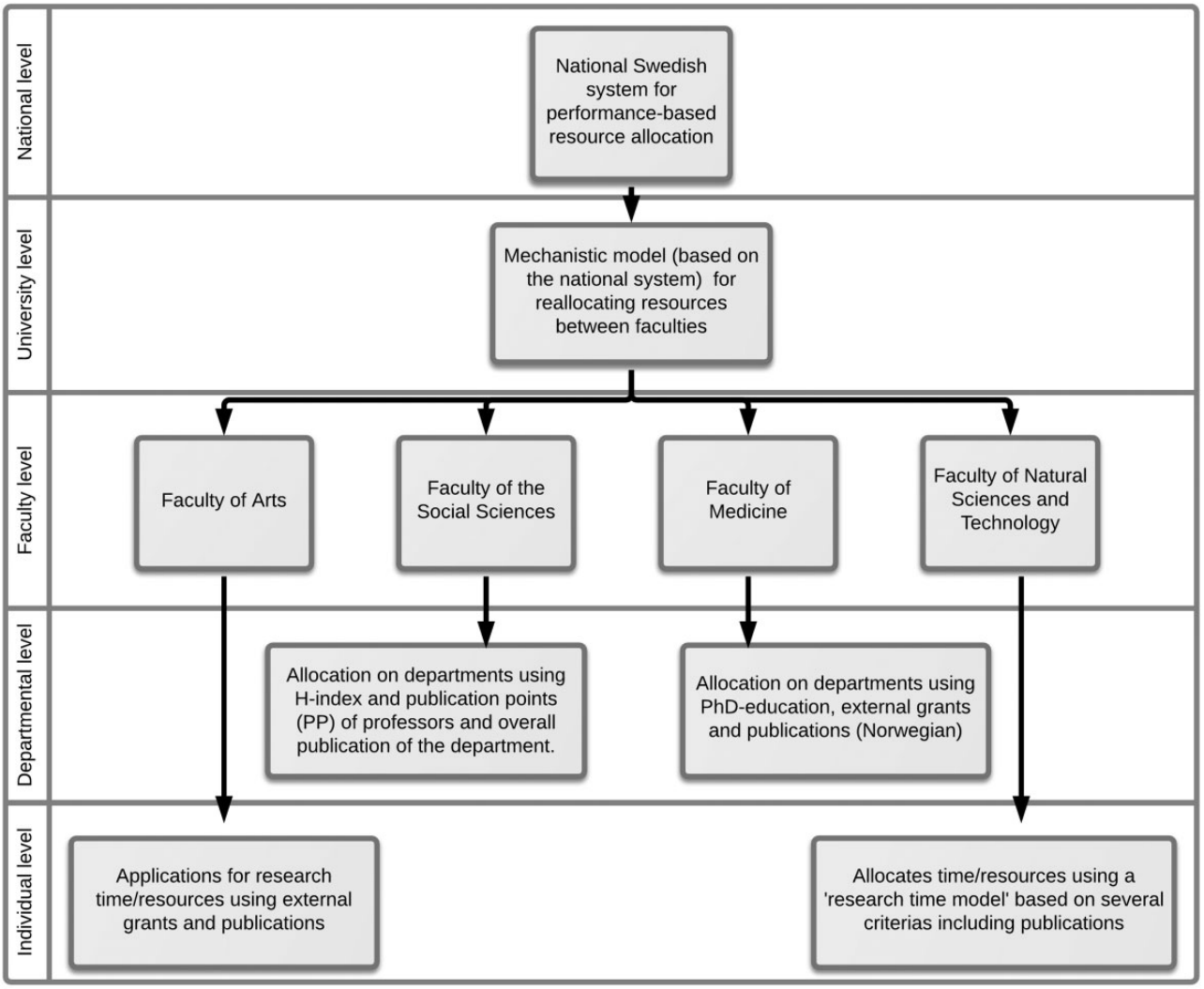

Figure 2. The use of bibliometric measures for resource allocation at Umeå University.

Notes. Illustration based on *Kvalitetsfrämjande tilldelning av resurser för forskning och forskarutbildning vid Umeå universitet* (2012), *Kvalitetsbaserad tilldelning av resurser för forskning och forskarutbildning vid Umeå universitet – mekanistisk modell 2014–2016* (2012), *Humanistiska fakultetens kriterier för tilldelning av kvalitetsbaserad resurs för forskning* (n.d), *Revidering av den samhällsvetenskapliga fakultetens resursfördelningssystem avseende forskning/forskarutbildning* (2012) and *Förslag till beslut om ny modell för resultatbaserad resursfördelning till forskning vid medicinska fakulteten, Umeå universitet* (n.d).

evaluation systems are bound to be complex, contradictory, and messy.

6. Evaluating allocation systems

The findings above provide an initial overview of the various systems that are currently used to allocate resources based on performance measures at Swedish universities. In the next step, these findings are assessed using our three main criteria: legitimacy and appropriateness, organizational and methodological stability, and transparency, feedback, and learning.

6.1 Legitimacy and appropriateness

Many decisions on implementing output-based allocation systems do not provide a distinct rationale for the design of the system. If given, these often point to structures and developments on the national level, and systems are rarely designed for meeting the university's own requirements. Such deliberations are revealed in the rationale provided for the model used at UU as well as the one used at Malmö University. In both these examples, the design of their respective performance-based resource allocation models is highly influenced by the national model. The following motivation is given in the description of the model used at UU:

> In order to increase incentives for quality and improving Uppsala University's outcomes in government, quality-based resource allocation, the Senate decided (28 September 2011) that 10% of the general funding of research and education at the graduate level should be redistributed between the research areas on an annual basis. [...] If the national resource allocation model changes, the university's distribution model should, as far as possible, be revised so that it coincides with the government model. (Modeller för fördelning av statsanslag från konsistoriet till områdesnämnderna vid Uppsala universitet, pp. 1–2)[1]

The rationale is thus, not to increase quality *per se*, but quality, as it is operationalized in the national model. The main goal for the HEIs is therefore to score better than their competitors:

> Thus, each university should be observant of how it stands in comparison to others in this regard. For an institution to receive a larger share of research funding than invested, it is not enough to reach a better result compared to the previous year. Rather, results must rise in relation to other universities. (Malmö Högskola—benchmarking av forskning 2012, Dnr. Mahr. 69-2012/439, pp. 1–2)[2]

In some cases, the motivations for implementing a model are straightforward. In their proposal for publication-based allocation at the faculties of Arts at GU, three main reasons for implementing the model are outlined: (1) too much of the research output is published locally and in Swedish, (2) there is little correlation between research time and productivity, and (3) the departments lack clear research profiles. (Fördelning av fakultetsmedel, 2011/196). These problems were identified in a comprehensive evaluation report (RED10), and the proposed solution was an allocation model based on the Norwegian system. It is rare that the aim of evaluation is so clearly defined; yet, we find that these three, internationalization, productivity-based allocation of research time, and specialization, are explicit or implicit goals in many systems. A fourth goal, that of 'excellent' or 'world class research', is often found in models using citation-based measures.

As in almost all evaluation systems, activities are reduced to a few quantifiable factors, where external grants, publications, and citations are the most common ones employed for assessing research at Swedish HEIs. Yet, we find that there are attempts to include more activities, such as reviewing for research councils or taking on the role as external examiner of PhD theses, although these more inclusive models have mainly been applied on the level of departments and faculties.

There is a clear risk of micro accountability being reinforced in these systems. The Swedish system for resource allocation across HEIs clearly stated that it should not be used on individual or departmental levels, but unsurprisingly, this is what happens. Micro accountability becomes even more explicit in systems, such as the ones at LTU, LNU, and BTH, where funds are allocated on the individual level. In these systems, researchers are rewarded financially for publications. Such rewards can be used for additional research time, travels, and other expenses.

Finally, it is evident that alternatives to evaluation and resource allocation systems based on bibliometrics have rarely been discussed up until very recently, when the Swedish Research Council published a report describing a national PRFS based on peer-review panels rather than quantitative performance indicators (Forskningsk valitetsutvärdering i Sverige 2014).

6.2 Organizational and methodological stability

The quality of data is a critical issue for all evaluation systems, and the legitimacy of any model is dependent on the data used. The models described above use a range of different data on publications, grants, and citations. Local publication databases, such as GUP (Gothenburg University Publications) or DiVA (Digital online archive), a database used for publication archiving by 36 Scandinavian Universities, are commonly used for extracting data on publications. Data originating from local publication databases have the advantage, at least potentially, of full coverage of all publications from all fields. However, the completeness and correctness of these databases cannot be guaranteed, as a majority of all records are self-reported and not subjected to sufficient examination and verification. Hence, the classification of publications as 'scholarly' is not always straightforward, and the definition of peer review differs across departments and authors (Verleysen et al. 2014). The self-reporting of data also opens up for duping the system, and it might be tempting, especially in cases were reported data are used to access individuals or small departments, to index non-peer-reviewed publications as peer reviewed.

Another common source is *Web of Science* that has its own liabilities and problems pertaining to quality, consistency, and coverage, and these issues might be particularly worrisome when it comes to non-English materials (cf. García-Pérez 2011). A further drawback in using data from Web of Science is the cost of obtaining rights to the database, and the expertise needed to handle citation scores in a rigorous and reliable manner.

Models using Web of Science data are limited to fields, particularly in the areas of natural science and medicine, where a considerable number of publications are indexed in the database. The coverage of the social sciences and the humanities is rarely high enough for evaluative purposes (Nederhof 2006). The adequacy of different data sources and models has resulted in diverse systems using a plethora of measures. Although it has often been pointed out that traditional bibliometric methods are less attuned to the research practices of the humanities and the social sciences, these observations are not always reflected in the choice of evaluation systems at specific departments or faculties. Hence, the only faculty at Stockholm University using bibliometric indicators is the faculty of arts, while one of the few faculties that refrains from using bibliometric measures at GU is the medical sciences.

The context in which evaluations are conduced is seldom discussed in literature on the topic of performance-based research evaluation, although the further use of bibliometrics in research libraries has gained some attention in recent years (Åström and Hansson 2013). Yet, it would be naïve to believe that the location of the bibliometric function as well as the professional role of the evaluator does not influence how the assessment is conducted and perceived by the researchers being evaluated. The research library together with central management are usually the hosts for the bibliometric function at many HEIs. The choice of the library as the host of the bibliometric function is logical and in line with the general development towards integrating the library into the publication process. Librarians are also usually in charge of the institutional repository used for bibliometric analysis. Yet, there might also be a '[. . .] danger that the library, being seen as a more active participant in research policy, now turns from being a service or support function at the university to becoming one with an auditing or monitoring function, passing judgment on scholars'. (Åström and Hansson 2013). The use of external expertise is another option employed, although local expertise for performing bibliometric analyses is usually lacking. However, the employment of consultants is more common when producing single evaluation reports rather than reoccurring assessments. A drawback in using consultants is their lack of in-depth knowledge of the evaluated activity, thus reducing the possibility for direct feedback.

6.3 Transparency, feedback and learning

Generally, local systems for the performance-based allocation of resources at the university level are poorly described, and this lack of transparency is further illustrated by the problems of finding adequate

descriptions of these systems. Documentation on the implementation and construction, as well as the results of evaluations, is often incomplete, difficult to access, or non-existent. Even identifying informants that could answer our questions was in some instances quite difficult. Occasionally, as in LnU and LU, descriptions of the model existed only as work documents and were therefore not publically accessible or hard to find. In other instances, as in the case of HB, the model is mentioned briefly and little further information is given.

Feedback, or even the possibility that the system can be used for providing valuable information to departments or individual scholars, is rarely mentioned in system descriptions. In this sense, they work more like management systems than as evaluative systems, despite the fact that the explicit goal in many cases is to change research and publication practices in ways that are perceived as beneficial to the university. Thus, it seems somewhat contradictory not to advertise the system and make it transparent if the goal is to actually change the manner in which research is conducted. On the other hand, if the rationale for implementing the system is to be on a par with other universities and to uphold an aura of accountability, then it makes perfect sense not to advertise the system. Another reason for not advertising the system could be to not provoke further debate; evaluation methods—especially those utilizing bibliometric indicators—have been subject of much discussion, especially among humanities scholars (Hammarfelt and de Rijcke 2015).

We have yet to encounter any discussion regarding the costs of evaluation: actual expenses for employees performing the assessment or the time it takes for researchers and administrators to report publications and other activities to systems are seldom mentioned. Hicks (2012) found that costs were rarely discussed in the development of PRFS on the national level either. This seems to be in accordance with a general trend where evaluation costs are seldom discussed or questioned (Dahler-Larsen 2012: 184–7). This might especially be the case for indirect costs, such as the recent work in adapting local databases to the national publication database of Sweden (*SwePub*), or the work done in registering and checking publications.

Few of the models reviewed in this study have been piloted or formally evaluated. One exception is the former system in UmU, which was assessed in an independent report (Utvärdering av kvalitetsbaserat [. . .] 2011) We do, however, suspect that test-runs might have been carried out informally. Overall, few mechanisms for feedback and learning were found in these systems, and discussions regarding their implementation and design are uncommon. Adequate and accessible documentation are also lacking in many cases, although a few exceptions provide thorough descriptions explaining the measures used (e.g. KI, UmU).

In all, we found little intent and few incentives to use these systems for learning or feedback among departments and scholars. Interaction with the system on the local level and from the perspective of the evaluand (i.e. departments and individual scholars) has primarily consisted of attempts at strategically maximizing their share of resources (gaming). Bibliometrics-based evaluation has rarely been used in strategic discussions, and if used, it is primarily in single evaluations, where bibliometric analyses and peer-review panels are combined.

7. Discussion and outlook

Our findings show that almost all major HEIs in Sweden use bibliometric measures for resource allocation; bibliometrics are applied at 24 of 26 HEIs in our sample, although the extent and the level differ considerably. Systematic bibliometric measurement can thus be considered to be an integrated part of everyday practices at Swedish universities. Due to the lack of comparable data, it is impossible to judge if this is unique for Sweden or if the use of bibliometric measures is as pervasive in other countries. The implementation of a national performance-based indicator appears to be one viable explanation for the increased use of bibliometric indicators across Swedish universities, and a recent study regarding the use of bibliometric indicators in Norway supports this conclusion (Aagaard 2015).

Evaluation systems are short-lived and ever-changing. The current Swedish system for allocating resources has been highly questioned since its introduction in 2009, and in December, 2014, the Swedish Research Council presented the outlines of a new system. This system will use peer-review assessment through panels, although bibliometrics will play an informative roll in fields where it is deemed applicable (natural science, medicine, technology). If the proposal, which mainly builds on the systems used in the UK (REF) and in Australia (Excellence in research for Australia), is implemented as intended, the first round of assessment will take place in 2018. The turn towards peer review as the main method for distributing resources will undoubtedly have consequences on the university level. Our study clearly shows how measures and incentives on a national level trickle down and influence decisions on lower levels. The actual effects of this new system are difficult to foresee, as national models play out very differently in local contexts. Many of them appear to strengthen incentives given in the national models, while other systems question or even negate them by using alternative indicators. However, a qualified assumption is that the current emphasis on bibliometric indicators in Swedish academia will decrease, and models directly aligned with the present system are likely to be revised. A follow-up to this study in a few years' time could reveal much about the fluctuating landscape of indicators and assessment models in general, and the consequences of the 'new model' in particular.

Motivations for implementing PRFS are often not articulated, and when a rationale is given, it frequently refers back to the national system for resource allocation. The introduction of a PRFS is rarely driven by the specific needs of the HEI in question, rather it would seem that evaluation systems built on performance indicators have become the norm. The idea of evaluation is ingrained in the academic system, and few question either the need for or the overarching goals of evaluation. In the tradition of new institutionalism (DiMaggio and Powell 1983), we could explain the current focus on bibliometric measurement as a result of isomorphism; universities adopt bibliometric evaluation models either because they imitate other institutions or because they function under the same constraints (the national model). Yet, even if many models resemble each other, it is also evident that all have their unique features, which partly contradicts this explanation. The variety of systems used is dependent on the diversity of HEIs in our study. The size, degree of specializations, and the overall organization of the HEI play a major role. Large, old, and diversified universities (UU, LU, GU) usually opt for mixed systems that are designed to evaluate specific domains, while large specialized HEIs (KTH, KI, SLU) might choose models attuned to their particular domain. Our study also shows that smaller regional universities (HB, HH, HIS) generally prefer models based on publication counts (variants of the Norwegian model). These systems have the advantage of being applicable across disciplines and are often relatively easy and cheap to implement. Hence, the decision of implementing a performance-based allocation system may very well be a consequence of external pressures and be

influenced by the need to be on a par with other institutions. Furthermore, the heterogeneous nature of indicators and models may be explained by the need to present the implementation of evaluation systems as a strategic and independent decision: 'In the age of autonomy, university management wants to present a unique project, and 'not just another evaluation'(Karlsson et al. 2014: 249).

Publication databases, citation indexes, assessment systems, and evaluation professionals are all part of a growing infrastructure surrounding the evaluation of research (Wouters 2014). The extended use of commercial citation databases such as Thomson Reuter's *Web of Science* and Elsevier's *SCOPUS* are perhaps the most notable examples of how the available infrastructure not only provides necessary data for evaluation but, to a large extent, also determines what can and should be measured. National publication databases, such as *SwePub* in Sweden and *CRISTIN* in Norway, also define what should be regarded as a measurable output. Here, we only have to point at allocation models (LU, BTH) where researchers are directly rewarded for publishing articles in journals indexed in Web of Science or CRISTIN. The influence of these systems can duly be questioned, as they severely reduce the many graduations associated with 'research quality' to a question of being indexed in a specific database.

Another often-voiced critique of these infrastructures is their consistency, where the quality of self-reported data is one crucial issue. Especially, the definition of peer review and the classification of genres can be particularly troublesome when different disciplines are compared. Methods and materials can also be criticized on the more fine-grained level of indicators. For example, substantial criticism questions the use of journal impact factors for accessing the impact of research (Seglen 1997; Alberts 2013). It is also notable that models specifically designed to be used on a national level, and where use on lower levels of aggregation is strongly discouraged, are still in use on faculty, departmental, and even individual levels.

In general, these models work without providing feedback to the researchers being evaluated. It could be argued that these systems functions mainly as assessment tools rather than as learning tools. Allocation models are enforced with the explicit goal of rewarding specific behaviour—to publish more and/or to publish in a specific channel—clarity. Therefore, it would seem that transparency in these systems is crucial for reaching the desired result. However, it is surprising just how invisible these systems are within Swedish HEIs and how little information about the indicators used is available. This raises doubt about the actual impact of these measures. Models applied on the higher level of departments and faculties might in many instances have little, direct influence on the individual researchers. Yet, we also know that indicators used on a higher level of aggregation might play out quite differently across departments within a university (Aagaard 2015), and the responsiveness to evaluation models could therefore differ greatly within a university. Systems that directly reward the individual researcher are notable exceptions, as the use of bibliometric indicators here has direct, and sometimes comprehensive, consequences for the possibility of conducting research.

The general lack of documentation, debate, and evaluation of performance-based evaluation systems at Swedish universities can actually be interpreted as a strategy for avoiding critique. On one hand, university management is sensitive to a general demand for evaluation across many types of activities, but on the other, they recognize that bibliometric evaluation systems are controversial and questioned within the academic community. The solution is then to implement systems that are little advertised and in which negligible resources are redistributed. Thus, in many cases, evaluation systems are foremost a 'ritual affair' (Dahler-Larsen 2012: 20), where the fact that evaluation takes place is more important than the result. However, when substantial resources are reallocated, most notably when bibliometric indicators are used on the individual level, these systems have considerable impact on working conditions and research priorities. The outcome of bibliometric evaluation might then determine if a new research project can be started, if it is possible to attend a conference or to purchase necessary equipment. Thus, their role as reallocating devices must not be forgotten, as this function positions them as administrative tools that are firmly incorporated in the larger structure of the university. This study has foremost focused on the evaluative aspects of bibliometric allocation models, but we also acknowledge that these systems play an important role in how universities are governed and administrated.

If researchers were to respond to the incentives in the models described above, we would see major changes in publication practices in some research field; yet, we do not envision such a development, as disciplinary traditions and strong professional identities are likely to counterweight the pressure of evaluations. There is also a potential tension between demands to publish in highly ranked journals and a further stress engendered by, for instance, research funding organizations to use open access publishing. Furthermore, the incentives of PRFS might not only collide with disciplinary traditions, they can also clash with other university goals. Hicks (2012: 259) suggests that performance-based evaluation using bibliometric measures might be in conflict with the promotion of applied research and interaction with industry. Moreover, it is likely that evaluations might have more profound effects on research practices in a longer perspective, as younger researchers appear to be somewhat more pressed by external evaluations (Hammarfelt and de Rijcke 2015).

An evaluation perspective on PRFS provides an in-depth critique that goes beyond the specific indicators used. Our operationalization of Dahler-Larsen's (2012b) framework for evaluating evaluation into three distinct aspects, *legitimacy and appropriateness*, *organizational and methodological stability, and transparency*, *feedback, and learning*, offers a possible framework for such an effort. This study can thus be seen as a contribution to a line of research that views research evaluation as a contextualized process, where techniques, infrastructures, epistemic considerations, and disciplinary cultures play a crucial role. For it is indeed surprising, that evaluation systems—as one of the distinctive components of late-modern organizations—are rarely systematically assessed themselves.

Funding

This work was supported by Svea Bredal Foundation; Swedish Research Council (grant number 2013-7368) and Riksbankens Jubileumsfond: The Swedish Foundation for the Social Sciences and Humanities) (grant number SGO14-1153:1).

Acknowledgements

The authors wish to thank two anonymous reviewers for their helpful and valuable comments on the first version of this manuscript.

Notes

1. Translated from Swedish: För att öka incitamenten för kvalitetsutveckling och förbättra Uppsala universitets utfall i regeringens kvalitetsbaserade resurstilldelning beslutade konsistoriet 28 September 2011 att 10% av basanslaget för forskning och

utbildning på forskarnivå ska omfördelas mellan vetenskapsområdena årligen. [. . .] Om den nationella resursfördelningsmodellen förändras ska en översyn av universitetets fördelningsmodell ske så att den i möjligaste mån följer regeringens modell. (Modeller för fördelning av statsanslag från konsistoriet till områdesnämnderna vid Uppsala universitet, pp. 1–2).

2. Translated from Swedish: Även om de nationella indikatorerna har mött viss kritik (särskilt applikationen av bibliometriskt index) utgör de en viktig del i beskrivningen av ett lärosätes forskningsverksamhet, inte minst därför att indikatorerna på ett normativt sätt omfördelar en viss del av landets forskningsresurser. Varje lärosäte bör därför vara observant på hur det står sig i jämförelse med andra i detta avseende. Om ett lärosäte ska få tillbaka en större andel av forskningsanslaget än den som satsats, räcker inte ett bättre resultat jämfört med föregående år, utan relativa andelar måste vinnas för att den indikatordrivna delen av anslaget ska höjas (Malmö Högskola – benchmarking av forskning 2012, Dnr. Mahr. 69-2012/439).

References

Aagaard, K. (2015). 'How Incentives Trickles Down: Local use of a National Bibliometric Indicator system', *Science and Public Policy*, 1–13. doi:10.1093/scipol/scu087

Ahlgren, P., Colliander, C. and Persson, O. (2012) 'Field Normalized Citation rates, Field Normalized Journal Impact and Norwegian Weights for Allocation of University Research Funds', *Scientometrics*, 92/3: 767–80.

Alberts, B. (2013) 'Impact Factor Distortions', *Science*, 340/6134: 787.

Åström, F. and Hansson, J. (2013) 'How Implementation of Bibliometric Practice Affects the Role of Academic Libraries', *Journal of Librarianship and information Science*, 45/4: 316–22.

Burrows, R. (2012) 'Living with the h-index? Metric Assemblages in the ContemporaryAcademy', *The Sociological Review*, 60/2: 355–72.

Butler, L. (2003) 'Explaining Australia's Increased Share of ISI Publications—the Effects of a Funding Formula Based on Publication Counts', *Research Policy*, 32/1: 143–55.

Butler, L. (2010) 'Impacts of Performance-based Research Funding Systems: a Review of the Concerns and the Evidence', In: Chapter 4 in OECD, *Performance-based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings*. Paris: OECD Publishing.

Dahler-Larsen, P. (2012a) *The Evaluation Society*. Stanford: Stanford University Press.

Dahler-Larsen, P. (2012b) 'Evaluation as a Situational or a Universal good? Why Evaluability Assessment for Evaluation systems is a Good Idea, What it Might Look Like in Practice, and Why it is not Fashionable', *Scandinavian Journal of Public Administration*, 16/3: 29–46.

De Rijcke, S., Wouters, P., Rushforth, A., Franssen, T. and Hammarfelt, B. (2015) 'Evaluation Practices and Effects of Indicator Use – A Literature Review'. *Research Evaluation*, doi: 10.1093/reseval/rvv038, in press.

DiMaggio, P. J. and Powell, W. W. (1983) 'The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields', *American Sociological Review*, 48/2: 147–60.

Furner, J. (2014) 'The Ethics of Evaluative Bibliometrics'. In Cronin, B. and Sugimoto, C. (eds), *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*. Cambridge, MA: MIT Press.

García-Pérez, M. (2011) 'Strange Attractors in the Web of Science Database', *Journal of Informetrics*, 5/1: 214–18.

Gingras, Y. (2014) 'Criteria for Evaluating Indicators'. In Cronin, B and Sugimoto, C. (eds), *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*. Cambridge, MA: MIT Press.

Hammarfelt, B. and de Rijcke, S. (2015) 'Accountability in Context: Effects of Research Evaluation Systems on Publication Practices, Disciplinary Norms and Individual Working Routines in the Faculty of Arts at Uppsala University'. *Research Evaluation*, 24: 63–77.

Hicks, D. (2012) 'Performance-based University Research Funding Systems', *Research Policy*, 41/2: 251–61.

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S. and Rafols, I. (2015). 'The Leiden Manifesto for Research Metrics', *Nature*, 520, 429–31.

Karlsson, S., Fogelberg, K., Kettis, Å., Lindgren, S., Sandoff, M. and Geschwind, L. (2014). 'Not Just Another Evaluation: a Comparative Study of Four Educational Quality Projects at Swedish Universities', *Tertiary Education and Management*, 20/3: 239–51.

Lorenz, C. (2012) 'If You're so Smart, Why are You Under Surveillance? Universities, Neoliberalism, and New Public Management', *Critical Inquiry*, 38/3: 599–629.

Nederhof, A. J. (2006) 'Bibliometric Monitoring of Research Performance in the Social Sciences and the Humanities: A Review', *Scientometrics*, 66/1: 81–100.

Nelhans, G. (2013) *Citeringens praktiker*. Diss. Göteborg: Göteborgs universitet. <https://gupea.ub.gu.se/handle/2077/33516> accessed 18 Feb 2015.

Ossenblok, T., Engels, T. and Sivertsen G. (2012) 'The representation of the social sciences and humanities in the Web of Sciencea comparison of publication patterns and incentive structures in Flanders and Norway (2005–9).' *Research Evaluation*, 21/4: 280–290.

Power, M. (1997) *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.

Sandström, U. and Sandström E. (2009) 'The Field Factor: Towards a Metric for Academic Institutions', *Research Evaluation*, 18/3: 243–50.

Schneider, J. W. (2009) 'An Outline of the Bibliometric Indicator Used for Performance-Based Funding of Research Institutions in Norway', *European Political Science*, 8/3: 364–78.

Seglen, P. O. (1997) 'Why the Impact Factor of Journals Should not be Used for Evaluating Research', *BMJ*, 314/7079: 498–502.

Sivertsen, G. (2008) 'Experiences with a Bibliometric Model for Performance Based Funding of Research Institutions', In Goriaz, J. and Schiebel, E. (eds) *Book of Abstracts, 10th International Science and Technology Indicators Conference*, pp. 126–128. 17–20 September 2008. Austria: University of Vienna.

Strathern, M. (2000) 'The Tyranny of Transparency', *British Educational Research Journal*, 26/3: 309–21.

Van Noorden, R. (2010) 'Metrics: A Profusion of Measures', *Nature*, 465/7300: 864–6.

Verleysen, F., Ghesquière, P. and Engels, T. (2014) 'The Objectives, Design and Selection Process of the Flemish Academic Bibliographic Database for the Social Sciences and Humanities (VABB-SHW)'. In: Blockmans, W., Engwall, L. and Weaire, D. (eds.) *Bibliometrics: Use and Abuse in the Review of Research Performance*. London: Portland Press.

Weingart, P. (2005) 'Impact of Bibliometrics upon the Science System: Inadvertent Consequences?', *Scientometrics*, 62/1: 117–31.

Whitley, R. (2007) 'Changing Governance of the Public Sciences' In: Whitley, R. and Gläser, J. (eds) Sociology of the Sciences *Yearbook*, pp. 3–27. Dordrecht: Springer.

Wildgaard, L., Schneider, J. W. and Larsen, B. (2014) 'A Review of the Characteristics of 108 Author-level Bibliometric Indicators' *Scientometrics*, 101/1: 125–58.

Woelert, P. and Yates, L. (2014) 'Too Little and Too Much Trust: Performance Measurement in Australian Higher Education', *Critical Studies in Higher Education*, 56/2: 175–189.

Wouters, P. (2014) 'The Citation: From Culture to Infrastructure'. In: Cronin, B. and Sugimoto, C. R. (eds.) *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Performance*, Vol. 22, pp. 47–66. Cambridge, MA: MIT Press.

Appendix A. Translated questionnaire

1. Are bibliometric models used for allocating resources at your institution? If so, how is the system designed?
2. Do you apply the Norwegian model or the Swedish national model (VR*)? Or are other models used?
3. Are the models changed/altered in anyway?
4. How are multi-authored publications handled? Are they fractionalized or are full counts used?
5. At which level is the model used? University, faculty, department, or individual level?
6. Is there any documentation available regarding the model and how it is used? Is it possible for us to get access to this material?
7. Have there been any internal debates regarding the use of bibliometric measures for research allocation?
8. Is the unit gathering data and performing the evaluation a part of the library? If not, where in the organization is this function located?

*Swedish Research Council

Appendix B. Documentation on performance-based allocation of research funds in Swedish academia

Användning av bibliometri som delparameter för tilldelning av forskningsmedel till institutioner vid Karolinska Institutet och forskning vid SLL, Karolinska Institutets Bibliometristyrgrupp på uppdrag av Forskningsstragiska kommittén (FSK) vid Karolinska Institutet/SLL, 2009.

Ett lyft för forskning och innovation, Prop. 2008/09:50, *2008*

Forskningskvalitetsutvärdering i Sverige – Fokus: Redovisning av ett regeringsuppdrag rörande modell för resursfördelning till universitet och högskolor innefattande sakkunniggranskning av forskningens kvalitet och relevans, Vetenskapsrådets rapporter, 2014.

Förslag till beslut om ny modell för resultatbaserad resursfördelning till forskning vid medicinska fakulteten, Umeå universitet, (n.d)

Humanistiska fakultetens kriterier för tilldelning av kvalitetsbaserad resurs för forskning, Umeå universitet, (n.d)

Inventering av bibliometrisk verksamhet vid svenska lärosäten. Håkan Carlsson Carlsson, H. and Hällgren, M 2008-09-23, Dnr: 61/08 J, 2008.

Kvalitetsfrämjande tilldelning av resurser för forskning och forskarutbildning vid Umeå universitet, Umeå universitet, Dnr: UmU 200-792-12, 2012

Kvalitetsbaserad tilldelning av resurser för forskning och forskarutbildning vid Umeå universitet – mekanistisk modell 2014–2016, Dnr: 200-792-12, 2012

Från department till doktorand: På vilka grunder fördelas de direkta anslagen för forskning. Görnerup, E. Svenskt Näringsliv, Januari 2013.

Fördelning av fakultetsmedel. Humanistiska fakultetsnämden, dnr. V 2011/196 Göteborgs Universitet, 2011-09-07.

Malmö Högskola – benchmarking av forskning Dnr. Mahr. 69-2012/4392012, 2012

Modell för fördelning av forskningsbidrag till forskningsmiljöer, Högskolan i Halmstad, Dnr. 69-2011-328. 2011.

Publiceringsstöd till vetenskapliga artiklar och konstnärliga produktioner, 2014, Rektorsbeslut nr 171-14, Luleå Tekniska Universitet.

Resurser för kvalitet, SOU 2007:81, 2007

Revidering av den samhällsvetenskapliga fakultetens resursfördelningssystem avseende forskning/forskarutbildning, Umeå universitet, Dnr 211-2535-12, 2012

Styrmodell vid Högskolan i Borås, 2011-09-22, Dnr. 697-10-21, 2011

Utvärdering av kvalitetsbaserat resurstilldelningssystem på humanistiska fakulteten, David Sjögren, Umeå Universitet, 2011.