Recognition and reward in the academy

Recognition and reward in the academy

Valuing publication oeuvres in biomedicine, economics and history

607

Björn Hammarfelt

Swedish School of Library and Information Science, University of Borås, Borås, Sweden Received 4 January 2017 Revised 20 April 2017 5 June 2017 Accepted 19 June 2017

Abstract

Purpose – The publication oeuvre of a researcher carries great value when academic careers are assessed, and being recognised as a successful candidate is usually equated with being a productive author. Yet, how publications are valued in the context of evaluating careers is so far an understudied topic. The paper aims to discuss these issues.

Design/methodology/approach – Through a content analysis of assessment reports in three disciplines – biomedicine, economics and history – this paper analyses how externalities are used to evaluate publication oeuvres. Externalities are defined as features such as reviews and bibliometric indicators, which can be assessed without evaluating the epistemological claims made in the actual text.

Findings – All three fields emphasise similar aspects when assessing: authorship, publication prestige, temporality of research, reputation within the field and boundary keeping. Yet, how these facets of quality are evaluated, and the means through which they are assessed differs between disciplines. Moreover, research fields orient themselves according to different temporal horizons, i.e. history looks to the past and economics to the future when research is evaluated.

Research limitations/implications – The complexities involved in the process of evaluating candidates are also reflected in the findings, and while the comparative approach taken effectively highlights domain specific differences it may also hide counter-narratives, and subtle intradisciplinary discussion on quality. Originality/value – This study offers a novel perspective on how publications are valued when assessing academic careers. Especially striking is how research across different fields is evaluated through different time horizons. This finding is significant in the debate on more overarching and formal systems of research evaluation.

Keywords Research quality, Research evaluation, Content analysis, Peer review, Reward system of science, Academic careers

Paper type Research paper

We appreciate the value and significance of a scholar only from his writings; he who does not write, seems unsuited to the world of scholarship today, and when someone applies for a position, one always asks what he has written (Johan Adam Bergk, Die Kunst, Bucher zu lezen: Nebst bemerkungen über Schriften und Schriftsteller, 1799, from Josephson, 2014).

Introduction

Reputation and recognition gained through publications has been a crucial merit for career advancement in academia since the birth of the research university in the late eighteenth century (Clark, 2008; Josephson, 2014). The ability to publish research is instrumental both

© Björn Hammarfelt. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at http://creativecommons.org/licences/by/4.0/legalcode

This study was supported by the Swedish Research Council (Grant No. 2013-7368) and The Swedish Foundation for the Social Sciences and Humanities (Grant No. SGO14-1153:1).



Aslib Journal of Information Management Vol. 69 No. 5, 2017 pp. 607-623 Emerald Publishing Limited EDO-3806 DOI 10.1108/AJIM-01-2017-0006 for gaining recognition within a specific field of research, and for the possibility of getting a permanent position at a university or a research institute. The reputation of an academic is dependent on their recognition among a wider community of peers, which means that the research field, rather than the institution, is the venue where careers are valued. In this sense, research fields are what Whitley (2000, p. 48f) calls "reputational work organisations" where labour market standing is determined by reputation among colleagues. Generally, it is assumed that the competition for positions in these reputational organisations has increased over the last decades, and while idioms like "publish or perish" are usually reiterated rather carelessly there appears to be some substance to the claim about increasing pressures to publish (Van Dalen and Henkens, 2012).

Academic researchers are continuously evaluated on the basis of their publication record, either as part of informal assessments or in the form of more regular systems of evaluation. A formal evaluation, which may have significant consequences for the individual career, takes place when applicants for an academic position are evaluated on the basis of their research merits, teaching and administrative skills. In this study, a selection of 45 assessment reports from four major universities in Sweden are used to study how publications are valued in this context. Commonly, the number and quality of publications are two main criteria through which research quality is evaluated. However, more exact studies of how research quality is defined in the context of evaluating candidates for academic positions are quite rare (Hemlin and Montgomery, 1993; Nilsson, 2009; Hammarfelt and Rushforth, in press), and research on conceptions of research quality has foremost been focussed on the peer review process of grants (see e.g. Langfeldt, 2001; Lamont, 2009; Van Arensbergen et al., 2014) rather than on academic careers. Moreover, the literature on academic careers tends to focus on structural aspects such as differences between national career systems (Musselin, 2009) or systematic discrimination based on gender (Steinpreis et al., 1999), while actual evaluation procedures have attracted less attention.

In focussing on how contextual information, such as information on the status of the publication channel, or externalities (e.g. bibliometric measures), are brought in to evaluate candidates this study engages in the current debate on peer review and indicator use in research assessment (Wouters et al., 2015). Externalities are defined as features such as publication channel, age of the texts, reviews, bibliometric indicators and prizes, which can be assessed without evaluating the epistemological claims made in the actual text. Recent research has shown how indicators are employed as "judgment devices" (Karpik, 2010) when evaluating research. The journal impact factor (JIF) has been identified as one frequently used such device which is integrated in the field of biomedicine where it also affects epistemological considerations (Rushforth and de Rijcke, 2015). The present study broadens the perspective introduced in these studies by engaging with contextual information about publications that might be used in similar ways, but which must not directly involve the use of bibliometric indicators. Thus, the purpose of this study is to provide a more detailed understanding of how "research quality" is defined and constructed in the context of evaluating the publication oeuvres of candidates for academic positions.

The analysis combines a theoretical framework for analysing field differences developed by Whitley (2000), and the theory of "judgment devices" formulated by Karpik (2010). Three fields of research – biomedicine, economics and history – were deliberately selected to highlight distinctive disciplinary valuation practices, although similarities in-between fields will also be emphasised. These fields were chosen on the basis of their being large high status fields both within and outside academia.

Biomedicine, a field in which principles of biology and chemistry are applied to clinical practices, consists of several subfields including molecular biology and biochemistry. The field has taken a central position in recent debates on the supposed crisis in science, where issues regarding how research and researchers are evaluated take a prominent position

(Benedictus *et al.*, 2016). However, more systematic studies, which go beyond anecdotal evidence on how research is evaluated in the recruitment of medical researchers, are still scarce (cf. Hammarfelt and Rushforth, in press).

Economics is one of the largest and most influential disciplines in the social sciences, and it is closely connected to the state and the economy at large (Maeße, 2017). The role of metrics and journal rankings has also been debated in the field of economics (Tourish and Willmott, 2015), but few studies have looked at how these measures are used when evaluating research.

History, on the other hand, is a discipline that is sometimes described as straddling the border between the social sciences and the humanities. The choice of history in this study is partly warranted by a current debate within the field where issues such as publishing preferences (Hammarfelt and de Rijcke, 2015), favoured publication language (Salö, 2016) and choice of dissertation form (Jezierski, 2016) are discussed.

Structure of the paper

First, a short overview of research on perceptions of scientific quality in general, and in the context of assessing individual researchers in particular, is presented. The subsequent section introduces the analytical frame developed by Whitley (2000), as well as the theory of judgment devices suggested by Karpik (2010). Material and methods are thereafter presented and the recruitment system in Swedish academia is briefly explained. This section is followed by the findings which are structured on five main themes identified in the material: authorship, publication prestige, temporality, reputation within the field, and boundary keeping. The concluding section summarises and discusses the implications of this study, while at the same time pointing to its limitations.

Scientific quality and the evaluation of careers

Conceptualisations of "scientific quality" in the context of peer review are a reoccurring topic in the literature. A noticeable strand within this area is studies looking at the work of grant panels, and how notions of quality are negotiated in this context. Seminal works, like Lamont's (2009) study of peer review, show how field-specific quality criteria are negotiated in multidisciplinary panels. Following in this tradition, several studies examine how judgments are made and negotiated in panels evaluating research grant applications (Langfeldt, 2001; Roumbanis, 2016). The present study distinguishes itself from these approaches in several ways: it concerns itself with intradisciplinary peer review; it looks at peer review that is done remotely (not in panels); and it uses reports, not interviews or ethnographic observation, as its primary material.

Conceptualisations of quality when evaluating and ranking candidates for academic positions have been much less studied, perhaps due to difficulties in gathering empirical material on procedures for evaluating candidates. In the literature we find two examples, both from Sweden, which have analysed how quality is defined in external assessment reports. Hemlin and Montgomery (1993) looked at assessment reports concerning candidates for 31 professorships in the humanities, the social sciences, medical sciences and natural sciences. They found considerable overlaps in how quality was judged across research fields, for example, mentions of methods, "problems" and "results" were frequent and "stringency" and "novelty" were deemed as important attributes for high quality research across all domains. The humanities and the social sciences stood out by highlighting attributes such as "reasoning" and "writing style" while "extra scientific relevance" and "international relations" were deemed as more important in the medical and natural sciences. Overall, Hemlin and Montgomery (1993) found that these differences could be explained by the division between "hard" and "soft" sciences, with the tentative conclusion that hard sciences are more easily evaluated due to greater agreement within the

field on theories and more "exact results". While this study was groundbreaking in its use of evaluation reports and therefore important for this paper, it also has limitations. First, the rather dated material (reports from 1981 to 1984) makes it less relevant in a contemporary context, and the use of very broad research areas (social sciences, humanities, natural sciences and medicine) – instead of disciplines – makes less sense if we agree with Whitley's assertion that research fields are the main contexts in which academic careers are evaluated.

The qualitative and comparative approach developed by Nilsson (2009) is of greater relevance for the present study. By studying assessment reports across three disciplines, physics, political science and literature, over a time period of 45 years Nilsson depicts how notions of quality have developed over time. Her approach of tracing conceptualisations of research quality using qualitative content analysis is a direct inspiration for this study. However, while she chose to select a few reports for each year, the present study gathers instead a larger number of contemporary reports in order to get a deeper understanding of how conceptualisations of quality are expressed when evaluating careers. The focus on the evaluation of careers, and publication oeuvres more specifically, as well as the emphasis on the use of contextual information, externalities and indicators, also signals distinctive differences in the approach adopted here compared to Hemlin and Montgomery (1993) and Nilsson (2009).

Hammarfelt and Rushforth (in press) analysed the use of metrics in assessment reports in biomedicine and economics. Their findings indicate that both disciplines use metrics rather extensively to assess candidates, but the type of use is dependent on the organisation of the field and on specific disciplinary publication patterns. The study also showed how bibliometric indicators are used as "judgments devices" to differentiate between candidates. The focus of the present study is more expansive as it incorporates a broader set of externalities used in the evaluation of the quality of publications.

Analysing evaluation reports

The choice of material and methodology used in this study is inspired by Hemlin and Montgomery (1993) and Nilsson (2009). While the former used an approach involving rather quantitative coding, the methodology adopted in the current study is best described as a qualitative content analysis where quotes, rather than statistics, are used to illustrate findings. In this sense, the current paper follows the path laid by Nilsson (2009) in her dissertation. Similarly to Nilsson I have chosen three fields – biomedicine, economics and history – which, to some extent, represent three "cultures" (social science, natural science and the humanities). Hence, the overall design of the study and the selection of fields assume that disciplinary differences might be a fruitful approach for studying how academic worth is judged. Yet, in order to avoid a simple confirmation of rather established conceptions of differences across disciplines special attention has been paid to details, which may contradict this neat separation of fields.

A total of 15 external assessment reports from each discipline were randomly selected from a larger collection of reports collected from four universities in Sweden (Lund University, Umeå University, University of Gothenburg and Uppsala University). A total of 45 reports, each comprising between 1-38 pages, was deemed large enough to provide a variety of different types of reports, while maintaining the possibility for a detailed analysis of the arguments made in each report[1]. Material from a ten-year period, 2005-2014, was collected. Although these are official documents that are accessible to anyone according to "offentlighetsprincipen" (principle of openness) it was decided to anonymise both referees and applicants. All reports were therefore coded based on year, field (biomedicine: bio, economics: eco and history: his) and university (Lund University: LU, University of Gothenburg: GU, Uppsala University: UU, Umeå University: UMU). Many of the reports, especially in economics and history, were written in Swedish or other Scandinavian

and reward in

the academy

languages and quotes used in the analysis were translated to English by the author (see Table AI for a full list of included reports).

The usual structure of these documents can be summarised as follows: first, a general introduction presenting the assignment, followed by detailed descriptions of each candidate and concluded with a ranking of applicants. Although there are small differences in the instructions for the external referees at each institution, these seem to have little influence on how the reports are written. Previous studies of peer review processes also suggest that formal instructions have little influence on actual evaluative procedures (Langfeldt, 2001, p. 837).

The common routine for recruiting academic personnel in Sweden is briefly described as follows: a decision to recruit is made by the head of the department or the dean; a description of the position and the qualifications needed to acquire the position is drafted and the job opening is advertised; applications from possible candidates, containing a CV, selected publications, and a description of pedagogical merits are submitted; external referees are chosen to access and sometimes even rank candidates; these assessments together with interviews and trial lectures by the leading candidates are used to form a final ranking of candidates (usually by a recruitment board); and based on this ranking the formal recruitment decision is made by the relevant authority (e.g. department head or dean).

My focus is specifically on stage 4 when CVs and a selected number of publications (usually around ten) are sent to external referees (so called "sakkunniga") who are assigned the task of making unbiased evaluations and ranking candidates. Reviewers usually make judgments on all merits, including teaching and administration, but research merits, and specifically publications, continue to play a key role in the final ranking (Brommesson *et al.*, 2016).

The number of candidates discussed in each text varied from 2 to 38 (reports assessing only one applicant were excluded). On average discussions of the qualities of each candidate ran over one page of text, but with large variances between candidates with top applicants being discussed more in-depth. Overall, the number of pages per document, and the number of pages per candidate are substantially smaller than what Hemlin and Montgomery (1993) found in their study in which each candidate was described on an average of 6.4 pages. However, their material consisted solely of assessment reports for professorships, while the current study also includes reports on candidates for lectureships. Still, it is evident that the overall length of reports and the number of pages per candidates have shortened substantially over the 30 years separating these studies.

The methodology chosen has similarities with directed content analysis, also called deductive content analysis (Hsieh and Shannon, 2005; Mayring, 2000) in that the analysis is guided by the theoretical frame provided by Whitley's theory on the organisation of research and Karpik's concept of "judgment devices". Initially, this theoretical viewpoint facilitated a focus on intellectual and social aspects of academic careers expressed through the evaluation of publication oeuvres using externalities. After a first reading of the documents five main themes, authorship, publication prestige, temporality, reputation within the field, and boundary keeping, were identified as the main evaluative categories. However, as will be evident in the material these categories are in no way mutually exclusive, and neat separations are not to be expected.

Theorising academic careers

Academic careers have characteristic features, which have to be considered when studying their evaluation. Gläser and Laudel (2015) find that two characteristics are distinctive: that the content of work (e.g. research done) plays an important role, and that the research community has a great deal of influence when evaluating academic careers. In view of these insights, Gläser and Laudel suggest that academic careers can be divided in three separate careers: a cognitive career, a community career and an organisational career. This separation accentuates how research fields and careers are both intellectual

Table I.Characterisation of research fields using Whitley's typology

(cognitive) and social (communal) in their nature. A similar approach is found in Whitley's (2000) study of the intellectual and social organisation of research fields. His theoretical framework for understanding the organisation of research fields is utilised for guiding and analysing the findings. In short, Whitley introduces two main axes that can be used to describe intellectual fields: mutual dependency and task uncertainty. Mutual dependency measures the degree to which a researcher is dependent on colleagues, while the degree of task uncertainty reflects agreement on the goals of research and the methods used. Whitley then continues by separating technical and task uncertainty and functional dependency and strategic dependency thus allowing for an intricate description of fields through 16 possible characterisations. How the three selected fields, biomedicine, economics and history, are depicted is summarised in Table I.

Whitley's theory lends itself well to more general discussions regarding disciplinary structure and its relation to evaluation practices. However, for a more detailed analysis, especially regarding the "externalities" used for evaluating publication oeuvres, more fine-grained analytical tools such as the concept of "judgment devices" are needed.

Iudgement devices

Field

When evaluating candidates referees face the task of assigning value to specific research accomplishments and to produce a ranking of applicants. This task is difficult because each academic career is distinctive and multidimensional. Such unique and not easily compared entities are termed "singularities" (Karpik, 2010). Examples of singularities are literary works or a medical doctor and when comparing and evaluating such "goods" consumers often make use of so called "judgment devices". Judgement devices provide external support for making and legitimating decisions, and their use in academic recruitment was first suggested by Musselin (2009). Musselin's study pointed to a more general use of judgment devices, but for the more detailed and comparative approach taken here it is important to consider the different types of devices identified by Karpik: appellations, cicerones, confluences, networks and rankings. Two of these, appellations and rankings, have previously been identified as particularly useful for understanding evaluation procedures in referee reports (Hammarfelt and Rushforth, in press). Appellations can be defined as a type of certification or brand; for example, prestigious journals or publishers that assign value to products (articles/books). Rankings, on the other hand, assign value by a hierarchisation of products based on specific criteria. Rankings can be further divided into "expert rankings" (e.g. prizes and diplomas awarded by juries) and "buyers rankings" (top ten products and bestseller lists) (Karpik, 2010, p. 46). A third judgment device, which is relevant for this study, is what Karpik calls cicerones, authorities in the form of guides or critics, which help consumers in making their choice. The use of judgment devices can be further understood in

Biomedicine	Professional adhocracy is characterised by combining reduced technical task uncertainty with
	high strategic task uncertainty. There is considerable standardisation of skills and technical
	procedures. No single group dominates when defining scientific criteria and various groups
	influence the field in terms of funding and employment
Economics	Partitioned bureaucracy combines high technical task uncertainty with low strategic
	uncertainty, and high strategic dependency. These are rule-governed fields, and hierarchically
	organised fields, where theoretical elaboration and analytical abilities carry greater value than
	empirical investigation
History	Fragmented adhocracy combines high task uncertainty with low degrees of mutual dependency.
-	In these fields research is personal and weakly coordinated, common-sense language is used when
	communicating results, and specialisation is formed around empirical objects

Characterisation according to Whitley (2000, p. 158f)

relation to the social and intellectual structure of research fields (Hammarfelt and Rushforth, in press), and field differences in the use of judgment devices is further elaborated upon in the discussion. First, however, it is necessary to present the themes which emerged from the analysis of the assessment reports.

Recognition and reward in the academy

Findings

The findings are structured around five main themes: authorship, publication prestige, temporality, reputation within the field, and boundary keeping. These themes emerged through an iterative categorisation of topics when analysing the reports. While this structure is useful for presenting the results in a systematised manner, it should be emphasised that such an arrangement is a simplification of a broader narrative. Moreover, many themes intermingle throughout the material and this is also visible in the analysis. As the current study has a focus on the evaluation of researchers as authors; it is logical to begin the analysis by scrutinising the notion of (co-)authorship across the three fields.

Authorship

It is well-established that notions and practices surrounding authorship differ considerably between research fields, which is reflected in that the average number of authors per publication varies from one or two in many humanities fields to tens- or even hundreds in the biomedical and the natural sciences (Marušić *et al.*, 2011). Naturally, these authorship practices have consequences for how collaboration in the form of joint publications is evaluated in the context of publication oeuvres. Moreover, research fields differ in their focus on either individual publications, or on the oeuvre as a whole. As Hemlin and Montgomery (1993) suggest, the medical and natural sciences tend to have a greater focus on the whole oeuvre, while the assessment of individual publications are the prime method through which research is assessed in the humanities.

Collaboration in the form of co-authorship is rarely touched upon in history, probably because it is quite rare, but there are instances when referees find it difficult to separate individual contributions and posit this as a potential problem: "[...] it is not always easy to separate the role and responsibility of the two authors" (His UU 2013, p. 3). However, on other occasions co-authorship might point to distinct qualities and due to its rarity it can be seen as a merit, rather than a problem: "[co-authorship][...] shows her ability to work and think together with other researchers and authors" (His LU 2011-1, p. 8). Overall, however, questions regarding co-authorship are few and co-authored pieces are uncommon.

The presence of several authors in the bylines is more frequent in economics, and typically two or three authors write the majority of papers, although examples of longer bylines are also found. In these instances of "multiauthorship", the value of a publication becomes unclear, as the role of the individual is hard to distinguish:

This resembles laboratory sciences where all those involved in a large project are included as authors. [...] The joint authorship makes it a bit hard to pinpoint individual contributions, but xxx's publication list includes several articles and papers written by him or with only a few co-authors, so clearly there is a fair amount of independent work (Eco GU 2007-3, p. 5).

What also matters is who you publish with, and papers published together with senior colleagues are generally viewed with a bit of scepticism: "As the other top candidates, xxx has a stellar publication record. However, it is a slight disadvantage that all his best papers are joint with senior co-authors" (Eco GU 2014-3, p. 7). Similar judgments are made in biomedicine, where too many publications with your former supervisors are seen as an indication of being too dependent: "She has not yet established herself as an independent researcher which is illustrated in that her former supervisor is co-author on 15/16 publications" (Bio GU 2006-1, p. 7). The author order, which has been found to play a central role for credit assignment in medicine

(Biagioli, 1998), is consistently referred to in the reports. Generally, it is first and foremost last authorships that are counted when publication oeuvres are valued, and being middle authors' counts for very little: "The results have been published in 41 multi-authored original publications, but most with the applicant in somewhat anonymous positions in the author sequences of the articles" (Bio LU 2011-1, p. 14). Prestige is instead attached to the first and the last position and the author order also signals degree of independence: "He clearly demonstrates independence with several publications as last or main author [...]" (Bio UU 2014-11, p. 1) and the last authorships also signify leadership: "He is frequently the senior author on his publications in recent years, indicating that he is clearly the leader behind the research line" (Bio LU 2005-6, p. 4). Hence, the ability to interpret author bylines, and gives credit based on this reading is a key competence when evaluating biomedicine, and the arrangement of authors as well as the reading of authorship order is highly standardised.

Hence, the reading and interpreting of author bylines is an established practice in biomedicine. The evaluation of multi-authored publications is less straightforward in economics as this quote illustrates: "It is always difficult to evaluate a candidate who publishes with many co-authors, especially when they are very senior" (Eco UU 2013-1, p. 4). In history, co-authorship it is still more of a curiosity rather than a problem, and the single author is the norm. Independence from senior researchers is also not an issue discussed in evaluating candidates, which is not unexpected given that research in history, according to Whitley (2000), is personal, weakly coordinated and highly specialised also early in the career.

Publication prestige

The type of publication channel that is assessed, and how it is valued varies considerably; monographs are the most prestigious publication channel in history, while journal articles are the most important merit in biomedicine and economics. Book chapters are not uncommon in economics, but in general they have less status than journal publications: "xxx has a series of articles in books about economic development but lacks scientific merits in the form of journal publications, which are needed to compete for the position" (Eco 2008-4, p. 2). Usually, evaluators in economics and biomedicine put considerable emphasis on publication channels, and papers in highly reputable journals are much valued. Publishing in more general high status journals is considered an important achievement in both fields, particularly in economics:

Xxx has maintained high productivity since the PhD defence in 1998, and has an impressive productivity. However, publications in more general journals would have helped to spread the results to other researchers (Eco GU 2008-4, p. 3).

Xxx shows relatively high productivity but his research has not yet reached the best journals (Eco GU 2008-4, p. 4).

Overall, the ability to publish in top journals emerges as the most important criteria for valuing careers in economics, and top journals, or highly ranked journals are mentioned in almost all reports. Sometimes it is a clearly distinctive factor: "I chose to rank first xxx because she is the only who has a top-5 publication [...]" (Eco UU 2013-1, p. 1). The decisive role of papers in top journals is explicitly commented upon by the reviewers, and in Eco GU 2012-5 the same phrase is repeated over and over again for eight candidates: "Furthermore it is recommended that xxx (to increase her relative competitiveness) improves his [sic] academic record by publishing in higher ranked journals". A similar view, but now in the context of evaluating a candidate's merits for professorship, is expressed by this reviewer:

A university that aims to compete at the first or second tiers in Europe should expect its full professors to show the ability to publish at least a few articles in the best journals in the field. Publishing a paper in a top finance journal requires a degree of effort, awareness of the latest thinking in the field, and excellence, which any number of articles in journals below second tier could not match (Eco UU 2006-1, p. 5).

and reward in

Apart from highlighting the significance of papers in top journals, as outlined above, these quotes also indicate the hierarchal structure of the field, where top institutions and top journals can easily be identified (Fourcade *et al.*, 2015). A logical consequence, as noted in the quote above, is that top researchers should publish in the best journals, and the highest-ranked universities should employ them. While hierarchies exist across all disciplines, it is probably warranted to claim that there is greater agreement on top journals or best universities in economics compared to many other fields. The hierarchal organisation of economics, which according to Maeße (2017) is further accentuated by the intertwined process of magnification and concentration, has direct consequences for how individual researchers are evaluated.

Top journals, or high impact journals, also have a distinct role in biomedicine, while other types of publications, including dissertations, matter less when evaluating researchers. Similarly to economics, reviewers of candidates for positions in biomedicine tend to discuss the status of the journal in which an article appears, and the names of prestigious journals, or in Karpik's terms, brands, to support their judgments:

For several years, he has published regularly as the corresponding author in excellent journals such as Chemistry and Biology, J. Biol. Chem, Blood, Biochemistry. He is also co-author of papers in prestigious journals such as Science and Nature (Bio UU 2008-1, p. 2).

The "market standing" of these "brands" are then often confirmed by the implicit and explicit use of the JIF:

He has published 27 papers and most of these are in high impact journals such as EMBO journal, Science, Journal of Clinical investigation, PNAS, JBC and Journal of Physiology (Bio LU 2005-6, p. 6).

Of the ten selected articles 4 have an impact factor (IP) > 4, 5 and 7 > 2 (Bio GU 2006-1, p. 2).

The JIF is here used as a judgment device that informs and supports assessment. Similarly to how journal rankings are employed in economics, JIF functions as a device which provides a shortcut to evaluating research, e.g. a paper published under the brand "Nature" automatically benefits from the reputation of the journal. Relating to Whitley's characterisation of biomedicine we can also regard the use of the JIF as a form of standardisation, which supports decision making in a situation where several different groups have to reach agreement when evaluating scientific quality.

Journal articles, especially if they are peer reviewed, are also a strong merit in the field of history and journals with good reputations also play a role here: "[...] a considerable number of her publications have appeared in renowned series or journals" (His GU 2014-1, p. 9). However, the skills associated with writing and publishing monographs are still highly valued: "The research is both in-depth and original, but its merits are devalued by the fact that xxx has not published any larger monographic work since the doctoral thesis in 1991" (His UU 2013-1, p. 3). The importance of the text's length is further accentuated by the use of the number of pages as one of the few "metrics" mentioned in evaluation reports in history:

The dissertation is long (622pp.) [...] The study is a large (579 pages) and is detailed research [...] (His GU 2014-1, p. 5).

Scientifically xxx is relatively well qualified with two monographs, and one longer article of 61 pages as well as a comprehensive report of 271 pages (His GU 2013-1, p. 17).

The use of numbers for measuring the length of publications is noteworthy as referees in history otherwise tends to rely on narrative accounts, which do not make use of quantitative data or metrics. Hence, the length of the publication is clearly an important factor when evaluating publications in history. Moreover, while the dissertation plays a minor, and in biomedicine, a negligible role when evaluating candidates, the assessment of

doctoral theses, almost exclusively in the form of a book, take up a considerable part of the evaluation report. In part this relates to the temporal horizons through which research is assessed.

Temporality

When reading the reports it becomes evident that the temporal foci of reviewers are quite distinctive in each discipline. As noted above, historians tend to spend a considerable time describing and valuing the dissertation, which in many instances is stated as being the candidates' strongest research merit. Many descriptions start out with a lengthy description of the dissertation work of the candidate, and the importance of the doctoral thesis is underlined: "xxx greatest scientific merit is his dissertation" (HIS GU 2007-1, p. 16), or in the case of a professor who is an author of several monographs: "The dissertation, which is of high scientific quality, is xxx strongest scientific qualification" (His UmU 2012-2, p. 5).

Dissertations are with no exceptions published as monographs, and many of them receive prizes, or other awards which are then mentioned as important merits. Hence, for younger researchers and even for more experienced scholars the dissertation is a persistent yardstick by which they are judged, and looking at the origin of an academic career will always be relevant. Particular emphasis is put not only on methods used or findings presented in the dissertation but also on language and presentation, thus similar to Hemlins and Montgomery (1993) aspect like writing style and reasoning are highlighted. In history, first impressions last – if not forever – for a very long time.

The dissertation plays a lesser role in economics and biomedicine, and here focus often lies on recent work. The dissertation in these fields is a starting point for a career, and rarely its high point. Evaluations of candidates in economics often go one-step further and evaluate research that has not been formerly published (e.g. pre-prints). Similar practices can also be found in biomedicine and history where drafts or book manuscripts under consideration are included in the evaluation. However, in economics forthcoming work is given greater weight compared to both history and biomedicine, and this difference can partly be explained by the tradition among economists to publish pre-prints ahead of formal acceptance. Yet, there are also suggestions that economics as a field is forward looking, and interested in being not only a descriptive but also a predictive science: "[...] [economist] 'live 'in the now', and see trajectories from the present forward', while sociologists have the reverse intellectual attitude, looking at the present as the outcome of a set of past processes" (Fourcade et al., 2015, p. 109, citing Abbott, 2005). The forward looking focus is reflected by many reviewers not only making judgments on research done, but also predicting which researchers have positive trajectories. This can in turn influence how researchers are compared:

As they have different expertise, it is hard to rank them. xxx and yyy have a richer publication record, but zzz is at an earlier stage in his career and on very positive trajectory (Eco GU 2014-3, p. 1).

Xxx has clearly improved his scientific qualifications over the last years, and there is reason to believe that he will publish well also in the future (Eco GU 2008-4, p. 8).

Career trajectories are also important in biomedicine, and successful publication careers are partly defined by how fast a candidate moves from being first author to last author. However, the sheer number of publications is, of course, also of great importance when evaluating careers: "His list of publications reveals a remarkable and unexplained decrease in scientific productivity during the last six years" (Bio LU 2011-1, p. 13). It is also apparent that publications are evaluated as part of an oeuvre, rather than as single works: "It is not only rarely seen, but also stimulating to evaluate such a consequential research career" (Bio LU 2011-1, p. 8). Overall, it is evident that these three disciplines employ slightly

and reward in

different temporal horizons when evaluating research. These can be schematically illustrated on a timeline (Figure 1).

Overall, historians are more inclined to value researchers based on older achievements and the first major work (the dissertation) is of great importance. Still, of course, career or in this case publication trajectories, also matter in history, as expressed by this reviewer: "His research does not show any clear progress" (His LU 2011-1, p. 13).

Overall, many of the evaluation reports build on an assumption of what might be defined as an "ideal trajectory" of the academic career. Thinking in terms of trajectories is a fundamental feature of western modernity (Appadurai, 2013 p. 223f), and this logic is apparent also when evaluating academic research (Felt, 2017). In this case, publications, (co-)authorship and indicators are used to position and compare individual careers against an "ideal trajectory"; a trajectory which is partly field specific. Yet, as shown in the next section, the type, amount and the temporal frequency of publication are not enough for evaluating a candidate; also reputation within the discipline is of importance when evaluating publication oeuvres.

Reputation within the field

The reputation that a publication of a scholar has gained within the discipline is an important criterion for assessing scientific merits. Often are external information, such as reviews, prizes, citations or similar, brought in to form and substantiate claims. As we will see different forms of "indicators" representing the reputation of a scholar are introduced depending on the discipline. These indicators are all said to represent the recognition and impact that a particular publication or an oeuvre has gained in the research community.

Prizes, peer review assignments, membership in associations and editorships are all important signs of recognition in history, and appreciation in form of reviews is quite often mentioned in connection to monographs. The finding that reviews play an important role for assessing reputation is in line with previous research suggesting that reviews might be seen as an indicator of impact (Zuccala and van Leeuwen, 2011). Prizes, often for dissertations and books, are also repeatedly used to present the reputation of a scholar. While national (Swedish) organisations are most visible we also see that international engagements in projects, review assignments and associations are highly valued. Candidates that exclusively publish for a Swedish audience are often criticised by reviewers, which might indicate that the criterion of "international reach" has gained in importance in comparison with Hemlin and Montgomery's (1993) study.

Prizes and book reviews serve in many ways the same role for historians as citations do in biomedicine and economics. These are used to showcase the recognition that particular publications have gained in the community:

The dissertation was awarded with the Geijerprize and is still her strongest merit (His GU 2013-1, p. 13).

Xxx has established herself as a leading researcher in her area. Which among other things is made visible in the reviews of her dissertation that have been published in international journals (His GU 2007-1, p. 16).



Figure 1. Schematic overview of temporal focus when evaluating research quality

Prizes can be seen as a type of endorsement, which in Karpik's vocabulary might be defined as an expert ranking, while the authority of reviews builds on the embodied and softer form of expertise in the form of critics or guides, or what Karpik (2010) terms cicerones.

In economics citations in specific publications, or to the whole oeuvre, are often used to measure the impact, and indirectly the reputation of researchers. For example, it can be stated that "[...] they have both made an impact on the profession, for instance both have a fair number of citations" (Eco GU 2008-5, p. 1), or similarly, it can be formulated in this way: "A search in Google scholar gives 197 hits which suggests an average/high visibility in the scientific community" (Eco UmU 2012-1, p. 1). Similar statements are made in biomedicine, with the difference that the number of citations per author and paper can be considerably higher than in economics: "His main author papers include papers with notably high citation rates (up to <1,000 citations), demonstrating his ability to publish visible cutting edge research" (Bio UU 2008-2, p. 2).

Overall, we find that a range of judgment devices are used across these fields, with significant overlap between them. However, it is important to note that the extent of use differs considerably between fields (Table II).

Prizes, for example, are rarely mentioned in biomedicine and economics (one instance each) but frequently used when evaluating careers in history. Similarly, it is also evident that these fields have distinct practices when it comes to defining and defending their borders.

Boundary keeping

External reports serve not only the purpose of assessing the merits of candidates, but these texts also make distinction between those that can be recognised as peers, and thus eligible for a position, and those that do not belong to the community. The disciplinary boundaries shield the market, and otherwise highly competent candidates have little chance to compete if they are deemed as "outsiders". Usually, reviewers refrain from making an assessment of such candidates "[...] scientific and pedagogic merits are primarily from the field of art history and he can therefore not be included on the shortlist" (His GU 2014-1, p. 11) or they make qualifications: "If his main and nearly exclusive research and publication area [...] is seen as belonging to the field of history, he would have a very strong and internationally qualified record, [...]" (His GU 2014-1, p. 15). Similar statements are also made in economics, "xxx is not an economist. All his publications are in non-economics journals" (Eco UU 2013-1. p. 5), or "The work shows good familiarity with the research area, but it is outside mainstream economics. This is shown also by the fact that xxx has no publications in general economics journals" (Eco GU 2007-3, p. 4).

Overall, it is evident that economists and historians are strict when it comes to upholding boundaries to other fields, but while publishing in key economic journals is enough for being recognised as a peer in economics, formal training as a PhD is a key qualification in history. This is probably due to relatively porous boundaries to other fields such as art history, economic history and history of ideas. The focus in biomedicine is more on specific competencies and whether the candidate will fit into a particular research profile or lab and,

Table II.
Judgment devices
used to assess the
recognition of
publications in the
discipline (type of
device according to
Karpik's typology)

	Biomedicine	Economics	History
Externalities			Reputation of publisher/journal (appellation) Prizes (expert ranking) Reviews (cicerones)

and reward in

the academy

as suggested by Whitley (2000, p. 161), a single group does not control the labour market in biomedicine. Using Whitley's theoretical frame it can be suggested that formal institutional origin – e.g. being trained as a historian – seems to play a decisive role in determining disciplinary borders in fields where agreement on research procedures or goals are less useful for defining the core of the discipline. Fields with a certain agreement on methods and procedures, might instead, as in the case of economics, define "membership" as having the skills needed to contribute to the advancement of the field.

Discussion

To evaluate and compare academic careers is a complex and demanding task, also for experienced reviewers. Careers, even when summarised in publication oeuvres, are multifaceted and not directly comparable. While disciplinary norms and "judgment devices" in the form of externalities may be of great help to reviewers, the many uncertainties and disagreements in the ranking of candidates are norm rather than the exception.

The complexities involved in the process of evaluating candidates are also reflected in the findings presented here, and while the comparative approach effectively highlights domain specific differences it might also hide counter-narratives and subtle intradisciplinary discussion on quality. With these limitations in mind, comparison also has distinct advantages in that it destabilises notions that "scientific" quality across fields and context can be distilled in a few quantifiable indicators. However, before returning to the wider policy implications that these findings might have, it may be worthwhile to concretise the main conclusions.

A first and overarching finding is that the three fields under study all emphasise similar aspects when evaluating candidates and these can be summarised in five themes: authorship, publication prestige, temporality of research, reputation within the field and boundary keeping. These aspects are also evident in the structure of all the reports, and a generic narrative form can be distilled from across all disciplines, making it accessible for practitioners that are familiar with the form but not experts on the evaluation procedures of specific fields.

While the criteria through which publications oeuvres are evaluated are fairly similar, the emphasis placed on these criteria varies greatly. Questions concerning co-authorship are prominent in biomedicine but less emphasised in economics. The reputations of publication channels in the form of highly ranked journals or journals with high impact matter a great deal in economics and biomedicine, while monographs and the length of publications are important for historians. Ways of assessing the impact of these publications in a community of peers also differs; citations are quite often utilised in biomedicine and economics, while prizes and book reviews are used as "indicators" of impact in history. Borders to other neighbouring disciplines are keenly defended in history and economics. Biomedicine is more porous. Overall, these results seem to support the notion that disciplinary differences do have great influence on evaluation procedures.

The evaluative procedures identified in these documents can then be further understood through Karpik's theory of judgment devices. On an abstract level, it seems that the dominance of appellations in the "standardized" field of biomedicine, rankings in the "hierarchically" organised discipline of economics, and the influence of cicerones in the "individualistic and weakly coordinated" field of history align well with the structure and organisation of these fields. However, it is worth emphasising that there also are several instances where the connection between disciplinary structure and evaluation procedures is less obvious, and judgment devices in the form of appellations and cicerones are found across all fields.

One feature, which is not easily incorporated in this arrangement, is how temporal aspects come to influence evaluation. It might in fact be argued that temporal dimensions

cut across all other dimensions, and that "trajectorial thinking" is an integral feature when evaluating research. This argument has recently been presented by Felt (2017), and it seems very appropriate in the context of studying how careers and publication oeuvres are valued. Indeed, the findings of this study indicate that research fields use distinct temporal horizons when evaluating research, which partly relates to epistemological factors. The ambition of economics to be a forward-looking field which tries to predict the future influences how research is evaluated, and the same applies to the field of history where past achievements, and especially the origins of academic careers are emphasised. Overall, time-perspectives seem to have a significant influence over how research is valued, yet temporality has so far been little discussed in the literature on research evaluation. A potentially fruitful direction for future research would be to look further into the issue of temporality and trajectoral thinking when evaluating researchers. More specifically such an approach could provide further knowledge on how aspects such as academic age and gender interrelate with ideas about an "ideal" trajectory of academic careers.

A common fallacy in recent debates on how to evaluate research is the assumption that agreement on the criteria for evaluating research also means that there is a general consensus on how these criteria should be applied. However, as this study has shown, the repertoire of indicators and externalities that are brought in to make and substantiate claims about the quality of research is distinctive for each field. The valuation of co-authorship or publication channels is field specific, as is the time horizon from which research is evaluated. Overarching systems for evaluating research employed by nations or institutions are by their very nature limited to using a very broad and crude set of indicators, and the measures used rarely reflect how scientific quality is defined within specific fields. The objective of this study is not to overcome this inherent tension between field-specific evaluation repertoires and more generic evaluation procedures. Rather, it illustrates that while a somewhat general agreement might exist on what constitutes research quality across fields, the actual tools and devices used to make these criteria tangible and comparable are distinct and not easily generalised.

The evaluation of applicants for academic positions based on their publication record is nothing new, and similarly we should not assume that different "short cuts", or judgment devices used for evaluating publication oeuvres is a late-modern innovation. As far back as the late eighteenth century concerns were expressed regarding the practice of over-emphasising opinions expressed in well-respected journals when evaluating candidates for academic positions (Josephson, 2014, p. 36). Similarly, it should be emphasised that the practice of reading texts and assigning scientific value to content, structure, style, findings and relevance of research is still an important, and in many cases the dominant form, of evaluation across all three fields. This kind of "classic", or perhaps idealised, peer review is, despite the availability of a range of indicators and metrics, still the primary practice used for evaluating candidates. So, in the context of evaluating candidates for academic positions it might be misleading to emphasise tensions between the use of indicators or other externalities, and "pure peer review". Rather, I suggest that the use of judgment devices should be seen as integrated within a larger set of evaluative practices. How these practices are formed in relation to disciplinary traditions, evaluative infrastructures and policy recommendations is therefore a question of great importance when assessing the quality of research.

Note

 The large difference in the total number of positions advertised at these universities over the period (2005-2014) and thus in available reports, 18 for history, 54 in economics and 132 in biomedicine, provided a further limitation to the number of reports that could be included.

Recognition

and reward in

References

- Abbott, A. (2005), "The idea of outcome in US sociology", in Steinmetz, G. (Ed.), The Politics of Methods in the Human Sciences: Positivism and Its Epistemological Others, Duke University Press, Durham, NC, pp. 393-426.
- Appadurai, A. (2013), The Future as a Cultural Fact: Essays on the Global Condition, Verso, London.
- Benedictus, R., Miedema, F. and Ferguson, M.W. (2016), "Fewer numbers, better science", Nature, Vol. 538 No. 7626, pp. 453-455.
- Biagioli, M. (1998), "The instability of authorship: credit and responsibility in contemporary biomedicine", *The FASEB Journal*, Vol. 12 No. 1, pp. 3-16.
- Brommesson, D., Erlingsson, G., Karlsson Schaffer, J., Ödalen, J. and Fogelgren, M. (2016), "Att möta den högre utbildningens utmaningar", IFAU, Uppsala.
- Clark, W. (2008), Academic Charisma and the Origins of the Research University, University of Chicago Press, Chicago, IL.
- Felt, U. (2017), "Under the shadow of time: where indicators and academic values meet", Special Issue: Engaging Science, Technology, and Society, Vol. 3, pp. 53-63.
- Fourcade, M., Ollion, E. and Algan, Y. (2015), "The superiority of economics", *The Journal of Economic Perspectives*, Vol. 29 No. 1, pp. 89-113.
- Gläser, J. and Laudel, G. (2015), "The three careers of an academic", discussion paper, Zentrum Technic und Gesellschaft, Technische Universitet Berlin.
- 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, Vol. 24 No. 1, pp. 63-77.
- Hammarfelt, B. and Rushforth, A.D. (in press), "Indicators as judgment devices: an empirical study of citizen bibliometrics in research evaluation", Research Evaluation, pp. 1-12, doi: 10.1093/reseval/rvx018.
- Hemlin, S. and Montgomery, H. (1993), "Peer judgements of scientific quality: a cross-disciplinary document analysis of professorship candidates", Science & Technology Studies, Vol. 28 No. 1, pp. 19-27.
- Hsieh, H.F. and Shannon, S.E. (2005), "Three approaches to qualitative content analysis", *Qualitative Health Research*, Vol. 15 No. 9, pp. 1277-1288.
- Jezierski, W. (2016), "Monografins elände: om svenska historikers avhandlingsformer", *Historisk tidskrift*, Vol. 2016 No. 136, pp. 232-242.
- Josephson, P. (2014), "The publication mill: the beginnings of publication history as an academic merit in German universities, 1750-1810", in Josephson, P., Karlsohn, T. and Östling, J. (Eds), The Humboldtian Tradition: Origins and Legacies, Brill Academic Publishers, Leiden, pp. 23-43.
- Karpik, L. (2010), Valuing the Unique: The Economics of Singularities, Princeton University Press, Princeton, NJ.
- Lamont, M. (2009), How Professors Think: Inside the Curious World of Academic Judgment, Harvard University Press, Cambridge, MA.
- Langfeldt, L. (2001), "The decision-making constraints and processes of grant peer review, and their effects on the review outcome", Social Studies of Science, Vol. 31 No. 6, pp. 820-841.
- Maeße, J. (2017), "The elitism dispositif: hierarchization, discourses of excellence and organizational change in European economics", *Higher Education*, Vol. 73 No. 6, pp. 909-927.
- Marušić, A., Bošnjak, L. and Jerončić, A. (2011), "A systematic review of research on the meaning, ethics and practices of authorship across scholarly disciplines", *Plos One*, Vol. 6 No. 9, pp. 1-17.
- Mayring, P. (2000), "Qualitative content analysis", Forum: Qualitative Social Research, Vol. 1 No. 2.
- Musselin, C. (2009), The Market for Academics, Routledge, London.
- Nilsson, R. (2009), God vetenskap. Hur forskares vetenskapsuppfattningar uttryckta i sakkunnigutlåtanden förändras i tre skilda discipliner, Acta Universitatis Gothoburgensis, Gothenburg.

- Roumbanis, L. (2016), "Academic judgments under uncertainty: a study of collective anchoring effects in Swedish research council panel groups", *Social Studies of Science*, Vol. 47 No. 1, pp. 95-116.
- Rushforth, A. and de Rijcke, S. (2015), "Accounting for impact? The journal impact factor and the making of biomedical research in the Netherlands", *Minerva*, Vol. 53 No. 2, pp. 117-139.
- Salö, L. (2016), Languages and Linguistic Exchanges in Swedish Academia: Practices, Processes, and Globalizing Markets, Department of Swedish Language and Multilingualism, Stockholm University, Stockholm.
- Steinpreis, R.E., Anders, K.A. and Ritzke, D. (1999), "The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: a national empirical study", *Sex Roles*, Vol. 41 Nos 7-8, pp. 509-528.
- Tourish, D. and Willmott, H. (2015), "In defiance of folly: journal rankings, mindless measures and the ABS guide", *Critical Perspectives on Accounting*, Vol. 26 No. 1, pp. 37-46.
- Van Arensbergen, P., van der Weijden, I. and van den Besselaar, P. (2014), "The selection of talent as a group process: a literature review on the social dynamics of decision making in grant panels", Research Evaluation, Vol. 23 No. 4, pp. 298-311.
- Van Dalen, H.P. and Henkens, K. (2012), "Intended and unintended consequences of a publish-or-perish culture: a worldwide survey", *Journal of the American Society for Information Science and Technology*, Vol. 63 No. 7, pp. 1282-1293.
- Whitley, R. (2000), The Intellectual and Social Organization of the Sciences, Oxford University Press, Oxford.
- Wouters, P., Thelwall, M., Kousha, K., Waltman, L., de Rijcke, S., Rushforth, A. and Franssen, T. (2015), "The metric tide", literature review, Supplementary Report I to the Independent Review of the Role of Metrics in Research Assessment and Management, HEFCE, London.
- Zuccala, A. and van Leeuwen, T. (2011), "Book reviews in humanities research evaluations", *Journal of the Association for Information Science and Technology*, Vol. 62 No. 10, pp. 1979-1991.

Appendix Recognition and reward in the academy

			the academy
Biomedicine	Economics	History	
Bio GU 2006-1	Eco GU 2007-3	His GU 2007-1	
Bio GU 2013-3	Eco GU 2007-5	His GU 2007-2	623
Bio GU 2013-9	Eco GU 2008-4	His GU 2009-2	020
Bio LU 2005-6	Eco GU 2008-5	His GU 2011-1	
Bio LU 2011-1	Eco GU 2012-1	His GU 2011-2	
Bio UU 2008-1	Eco GU 2012-5	His GU 2013-1	
Bio UU 2008-2	Eco GU 2014-1	His GU 2014-1	
Bio UU 2010-3	Eco GU 2014-3	His LU 2011-1	
Bio UU 2012-7	Eco LU 2011-3	His LU 2011-2	
Bio UU 2012-10	Eco LU 2011-5	His LU 2013-2	
Bio UU 2014-3	Eco UmU 2012-1	His UmU 2010-1	
Bio UU 2014-4	Eco UU 2006-1	His UmU 2010-2	
Bio UU 2014-5	Eco UU 2009-3	His UmU 2012-2	
Bio UU 2014-6	Eco UU 2012-1	His UU 2013-1	Table AI.
Bio UU 2014-11	Eco UU 2013-1	His UU 2013-2	Analysed reports

About the author

Björn Hammarfelt, PhD, is a Senior Lecturer at the Swedish School of Library and Information Science, University of Borås, Sweden and a Visiting Scholar at the Centre for Science and Technology Studies (CWTS), Leiden University. His research is situated at the intersection between information science and sociology of science, with a focus on the organisation, communication and evaluation of research. Björn Hammarfelt can be contacted at: bjorn.hammarfelt@hb.se