15

Assessing Academic Careers: The Peer Review of Professorial Candidates

Björn Hammarfelt

Introduction¹

Academic careers have two characteristic features, which have to be considered when studying their evaluation: the content of work (e.g. research done) plays an important role, and the research community has a great deal of influence when evaluating academic careers. The dependence on colleagues means that the reputation of an academic is dependent on their recognition among a wider community of peers. The primary means for gaining a reputation among colleagues is through publications, and the recognition of a researcher is largely dependent on their writings. In fact, reputation and recognition gained through publications has been a

¹ This chapter is a revised, updated and shortened version of an article published in *ASLIB Journal of Information Management* under the title: “Recognition and reward in the academy: Valuing publication oeuvres in biomedicine, economics and history” (2017). <https://doi.org/10.1108/AJIM-01-2017-0006>

B. Hammarfelt (✉)
University of Borås, Borås, Sweden
e-mail: bjorn.hammarfelt@hb.se

crucial merit for career advancement in academia since the birth of the research university in the late eighteenth century (Josephson, 2014). Generally, it is assumed that the competition for positions in academia 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 the increasing pressure to publish (Van Dalen & 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. This chapter looks at discipline specific evaluation practices in three fields; biomedicine, economics and history. The material consists of reports (*sakkunnigutlåtanden*) commissioned by Swedish universities when hiring new professors. Independent referees hired to evaluate and compare candidates author these texts. The approach here is not so much to study what constitutes ‘value’ in these evaluations, rather the focus is on how ‘value’ is enacted with special attention to the kind of tools—judgements, indicators and metrics—that are used. 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 & Montgomery, 1993; Nilsson, 2009; Hammarfelt & Rushforth, 2017), and research on conceptions of research quality has foremost been focused 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 focusing 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 ‘judgement devices’ (Karpik, 2010; Hammarfelt & Rushforth, 2017) when evaluating research. The journal impact factor has been identified as one frequently used such device which is integrated in the field of biomedicine where it also affects epistemological considerations (Rushforth & 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.

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.

The chapter starts with a short outline of research on peer review and perceptions of scientific quality. The subsequent section introduces the theory of judgement devices suggested by Karpik (2010), and the analytical frame developed by Whitley (2000). Material and methods are thereafter presented and the recruitment system in Swedish academia is briefly explained. The findings 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.

Picking the Best: Peer Review in Assessment Procedures

Conceptualisation of 'scientific quality' in the context of peer review is a reoccurring topic in the literature. A noticeable strand within this area is studies that look 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 judgements are made and negotiated in panels evaluating research grant applications (Langfeldt, 2001; Roumbanis, 2017). The present study distinguishes itself from these approaches in several ways; it concerns itself with intra-disciplinary 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 research quality when evaluating and ranking candidates for academic positions has been much less studied, perhaps due to difficulties in gathering empirical material on procedures for evaluating candidates. 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. Yet they also found differences which could be explained by the division between 'hard' and 'soft' sciences.

The qualitative and comparative approach developed by Nilsson (2009) is of 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. 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. A similar approach, but with a focus on teaching merits, is Levander's (2017) study of how pedagogical merits are evaluated. A notable finding in this study is that research

merits—often in the form of publications—usually have greater impact on the final ranking compared to other accomplishments.

Hammarfelt and Rushforth (2017) 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 showed how bibliometric indicators are used as ‘judgements 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 Referee Reports

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. Three fields—biomedicine, economics, and history—which, to some extent, represent three ‘cultures’ (social science, natural science and the humanities) have been selected for analysis. 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.

Fifteen external referee reports from each discipline were randomly selected from four universities in Sweden (Lund University, Umeå University, University of Gothenburg and Uppsala University). A total of 45 reports, each comprising about 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.²

²The number of available reports at these universities over the period (2005–2014), 18 for history, 54 in economics and 132 in biomedicine, provided a further limitation to the number of reports that could be included in this study.

Material from a ten-year period, 2005–2014 was collected. Although these are official documents that are accessible to anyone according to ‘offentlighetsprincipen’ (the principle of public access to public records), 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 languages and quotes used in the analysis were translated to English by the author.

The common routine for recruiting academic personnel in Sweden can briefly be described in six steps: (1) a decision to recruit is made by the head of the department or the dean, (2) a description of the position and the qualifications needed to acquire the position is drafted and the job opening is advertised, (3) applications from possible candidates, containing a CV, selected publications, and a description of pedagogical merits are submitted, (4) external referees, are chosen to assess and sometimes even rank candidates, (5) 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), (6) 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 judgements on all merits, including teaching and administration, but research merits, and specifically publications, continue to play a key role in the final ranking (Levander, 2017). 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.

The methodology chosen has similarities with directed content analysis, also called deductive content analysis (Hsieh & 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 ‘judgement 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.

Judgement Devices and Intellectual Structure

When evaluating candidates, referees face the task of assigning value to specific research accomplishments and producing 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 judgement 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 judgement 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 are less applicable in the context of academic valuation; *networks* describe how the personal network of a buyer (friend and family) influences their choices, and *confluences* relate to who buyers navigate in a physical space, for example in a store. *Appellations* and *rankings* have previously been identified as particularly useful for understanding evaluation procedures in referee reports (Hammarfelt & Rushforth, 2017). *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 judgement 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 judgement devices can be further understood in relation to the social and intellectual structure of research fields (Hammarfelt & Rushforth, 2017). Thus, Whitley’s (2000) study of the intellectual and social organisation of research fields is used in order to analyse how judgement devices are employed in specific disciplinary contexts. How the three selected fields, biomedicine, economics and history, are depicted in Whitley’s framework is summarised below (Table 15.1).

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. Whitley’s theories are here used for the specific purpose of providing an analytical lens through which disciplinary differences in assessing careers become visible.

Table 15.1 Characterisation of research fields using Whitley’s typology

Field	Characterisation according to Whitley (2000, p. 158f.)
Biomedicine	<i>Professional adhocracy</i> 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	<i>Partitioned bureaucracy</i> 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	<i>Fragmented adhocracy</i> 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.

Merits and Their Assessment: Five Main Themes

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, or the Reading of By-Lines

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 either on 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 infrequent, 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 2013-1, 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 by-lines is more frequent in economics, and typically, two or three authors write the majority of papers, although there are instances of longer by-lines. In these instances of ‘multi-authorship’, 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 authorships make 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)

Who you publish with matters, and papers co-authored 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 judgements 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 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 by-lines, and give 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 is still more of a curiosity rather than a problem, and the single author is the norm. Independence from senior researchers is not an issue discussed in evaluating candidates, which is expected given that research in history, according to Whitley (2000), is personal, weakly co-ordinated and highly specialised even early in the career.

Publication Prestige and the Importance of Articles in ‘Top-Journals’

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). A similar view 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)

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. As argued by Hylmö (2018, p. 295) these top journals “merge with an understanding of something like a disciplinary core”, and in order to be accepted researchers need to adapt to an “established disciplinary style of reasoning”. 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 Mæße (2017), is further accentuated by resources and academic capital being concentrated to a few ‘top’ institutions, has direct consequences for how individual researchers are evaluated.

Top journals, or high impact journals, have a distinct role in biomedicine, while other types of publications, including dissertations matter less when evaluating research. Similar 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 judgements:

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 Journal Impact Factor (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)

The JIF is used here as a judgement 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; for example, 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 a strong merit in the field of history and journals with good reputations are appreciated: "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 is 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 (622 pp.) [...] 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 tend 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. The Importance of a 'Positive Trajectory'.

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 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). 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).

Dissertations are, with no exceptions, published as monographs, and many of them receive prizes, or other awards, which are 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 on not only methods used or findings presented in the dissertation, but also on language and presentation. Thus

similar to views of Hemlin and Montgomery (1993), aspects 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 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 judgements 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)

In biomedicine, successful publication careers are partly defined by how fast a candidate moves from being first author to last author. Moreover, the number of publications is of great importance for evaluating the direction of the career trajectory: “His list of publications reveals a remarkable and unexplained decrease in scientific productivity during

Fig. 15.1 Schematic overview of temporal focus when evaluating research quality

the last six years” (Bio LU 2011-1, p. 13). Here 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 different temporal horizons when evaluating research. These can be schematically illustrated on a timeline (Fig. 15.1).

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; Hammarfelt et al., 2020). 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 is not enough for evaluating a candidate, as reputation within the discipline carries great weight when evaluating publication oeuvres.

Racing for the Prize: Reputation Through Awards and Citations

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 & van Leeuwen, 2011). Prizes, often for dissertations and books, are 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)

Prizes can be seen as 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, “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

Table 15.2 Judgement devices used to assess the recognition of publications in the discipline (*type of device according to Karpiks typology*)

	Biomedicine	Economics	History
Externalities	Reputation of journal (<i>appellation</i>)	Reputation of publisher/journal (<i>appellation</i>)	Reputation of publisher/journal (<i>appellation</i>)
	Impact factor (<i>appellation</i>)	Impact factor (<i>appellation</i>)	Prizes (<i>expert ranking</i>)
	h-index (<i>buyers ranking</i>)	Citations (<i>buyers rankings</i>)	Reviews (<i>cicerone</i>)
	Citations (<i>buyers ranking</i>)	h-index (<i>buyers rankings</i>)	
	Prizes (<i>expert ranking</i>)	Prizes (<i>expert ranking</i>)	

in the scientific community” (Eco UmU 2012-1, p. 1). Similar statements are made in biomedicine, with the difference that the amount 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 <1000 citations), demonstrating his ability to publish visible cutting edge research” (Bio UU 2008-2, p. 2).

Overall, we find that a range of judgement devices is 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 15.2).

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

Boundary Keeping and the Shielding of Academic Markets

External reports serves 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 (cf. Levander, Forsberg, Lindblad & Bjurhammer, *this volume*). 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 short-list.” (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 economist 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, 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—for example, 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

The peer review of 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 ‘judgement 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. Importantly, reviewers not only have to be experts on current research in their field, but they must also be well acquainted with current assessment procedures and evaluation criteria’s used. As expressed by Hammarfelt and Rushforth (2017, p. 178), “it is knowing *how* and *when* to deploy indicators which should be considered the marker of expertise in such evaluative contexts.”

Still, 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 with in the field and boundary keeping. These aspects are evidently 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 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 judgement devices. On an abstract level, it seems that the dominance of *appellations* in the ‘standardised’ field of biomedicine, *rankings* in the ‘hierarchically’ organised discipline of economics, and the influence of *cicerones* in the ‘individualistic and weakly co-ordinated’ 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 judgement 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. 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 (Hammarfelt et al., 2020), yet temporality has so far been little discussed in the literature on research evaluation.

A common fallacy in recent debates on how to evaluate research is the assumption that agreement on the criteria for evaluating research 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 peer review 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 judgement 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, the use of judgement devices, for example bibliometric indicators, should be seen as integrated within a larger set of evaluative practices.

References

- Appadurai, A. (2013). *The future as a cultural fact: Essays on the global condition*. Verso.
- Abbott, A. (2005). The idea of outcome in US sociology. In G. Steinmetz (Ed.), *The politics of methods in the human sciences: Positivism and its epistemological others* (pp. 393–426). Durham, NC: Duke University Press.
- Biagioli, M. (1998). The instability of authorship: Credit and responsibility in contemporary biomedicine. *The FASEB Journal*, 12(1), 3–16.
- Felt, U. (2017). Under the shadow of time: Where indicators and academic values meet. *Engaging Science, Technology, and Society*, 3, 53–63.
- Fourcade, M., Ollion, E., & Algan, Y. (2015). The superiority of economics. *The Journal of Economic Perspectives*, 29(1), 89–113.
- Hammarfelt, B., & Rushforth, A. D. (2017). Indicators as judgment devices: An empirical study of citizen bibliometrics in research evaluation. *Research Evaluation*, 26(3), 169–180.
- Hammarfelt, B., Rushforth, A. D., & de Rijcke, S. (2020). Temporality in academic evaluation: ‘Trajectorial Thinking’ in the assessment of biomedical researchers. *Valuation Studies*, 7(1), 33–63.

- Hemlin, S., & Montgomery, H. (1993). Peer judgements of scientific quality: A cross-disciplinary document analysis of professorship candidates. *Science & Technology Studies*, 28(1), 19–27.
- Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288.
- Hylmö, A. (2018). *Disciplined reasoning: Styles of reasoning and the mainstream-heterodoxy divide in Swedish economics*. Diss. Department of Sociology, Lund University.
- Josephson, P. (2014). The publication mill: The beginnings of publication history as an academic merit in German universities, 1750–1810. In P. Josephson, T. Karlsohn, & J. Östling (Eds.), *The Humboldtian tradition: Origins and legacies* (pp. 23–43). Brill Academic Publishers.
- Karpik, L. (2010). *Valuing the unique: The economics of singularities*. Princeton University Press.
- Lamont, M. (2009). *How professors think: Inside the curious world of academic judgment*. Harvard University Press.
- Langfeldt, L. (2001). The decision-making constraints and processes of grant peer review, and their effects on the review outcome. *Social Studies of Science*, 31(6), 820–841.
- Levander, S. (2017). Den pedagogiska skickligheten och akademins väktare: Kollegial bedömning vid rekrytering av universitetslärare. Diss. Acta Universitatis Upsaliensis, Uppsala.
- Maeße, J. (2017). The elitism dispositif: hierarchization, discourses of excellence and organizational change in European economics. *Higher Education*, 6, 909–927.
- Marušić, A., Bošnjak, L., & Jerončić, A. (2011). A systematic review of research on the meaning, ethics and practices of authorship across scholarly disciplines. *Plos One*, 6(9), 1–17.
- Mayring, P. (2000). Qualitative content analysis. *Forum: Qualitative Social Research*, 1(2), Art. 20.
- Musselin, C. (2009). *The market for academics*. Routledge.
- Nilsson, R. (2009). God vetenskap. Hur forskares vetenskapsuppfattningar uttryckta i sakkunnigutlåtanden förändras i tre skilda discipliner. Diss. Acta Universitatis Gothoburgensis, Gothenburg.
- Roumbanis, L. (2017). Academic judgments under uncertainty: A study of collective anchoring effects in Swedish Research Council panel groups. *Social Studies of Science*, 47(1), 95–116.
- Rushforth, A., & Rijcke, S. (2015). Accounting for impact? The journal impact factor and the making of biomedical research in the Netherlands. *Minerva*, 53(2), 117–139.

- Steinpreis, R. E., Anders, K. A., & 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*, 41(7–8), 509–528.
- Van Arensbergen, P., van der Weijden, I., & 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*, 23(4), 298–311.
- Van Dalen, H. P., & 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*, 63(7), 1282–1293.
- Whitley, R. (2000). *The intellectual and social organization of the sciences*. Oxford University Press.
- Wouters, P., Thelwall, M., Kousha, K., Waltman, L., de Rijcke, S., Rushforth, A., & 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.
- Zuccala, A., & van Leeuwen, T. (2011). Book reviews in humanities research evaluations. *Journal of the Association for Information Science and Technology*, 62(10), 1979–1991.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.