

Safeguarding the unknown?

Quality of research in the performance measurement era at universities

David S. Bedford
University of Technology Sydney

Markus Granlund
University of Turku

Kari Lukka
University of Turku

Abstract

In this study we examine the practical meaning and employment of the notion of research quality in the academe. This study is inspired by a worry that the difficult-to-define notion of quality in research is potentially getting too simplistically determined by its measurable proxies, and whether academics, especially manager-academics, realise this risk, and how they deal with it. While previous studies provide relatively good visibility to the landscape of performance measurement in the university sector, we know little about how performance measurement systems (PMS) are mobilized locally, especially as comes to how one of the fundamental virtues of scientific work, that of quality, is perceived and managed. To examine these matters empirically, we conducted a comparative case study of two university faculties in a European country. Despite differences in the local PMS, manager-academics are found to have rather similar understandings of the meaning of quality in research. However, there were variations in the willingness and perceived need to exert their agency regarding how quality is operationalised. This is seen to be partly a function of how restrictive the local PMS is in terms of what constitutes desired academic performance, and the degree to which the PMS is relied upon to make judgmental evaluations of research quality. We conclude by commenting on how forces both outside and within universities are driving a more narrow understanding of what quality in research means in practice.

Introduction

‘Quality’ is a term that we often encounter in the academe. It is a taken-for-granted good and celebrated aspect of academic work overall – and not least so with regards to scholarly research. Quality is often discussed as if everyone knew what it means, but this is a far from unproblematic assumption, as the very concept of quality is challenging to even define. This difficulty is well illustrated in Robert Pirsig’s “Zen and the Motorcycle Maintenance” (1974), in which the key character, named Phaedrus, made the definition of quality a very serious issue and led him, after a most profound philosophical analysis, to conclude that a ‘true quality’ is undefinable. Whether this conclusion is inevitable is not central to our analysis, but rather how Pirsig (1974) manages to demonstrate how puzzling only defining the notion of quality can be viewed. In terms of our theoretical orientation, Pirsig (1974) provides a teasing ‘way of thinking’ about quality. It forms a starting point for our analysis, which – rather than making any further attempt to define the notion as such – elaborates this problematic and ambiguous concept by examining how it is rendered practical, that is, how it is understood and employed in practice in the current circumstances of research in the academe.

In broad terms, the theoretical domain of our study is *performance management* and its effects, especially in highly knowledge-intensive contexts, such as universities. By university performance management we refer to the complex including the increasing publish-or-perish culture and managerialism, which nowadays seem to go hand in hand in universities all over the world (e.g., Neumann & Guthrie, 2002; Thornton et al., 2012; Townley, 1997; Alvesson et al., 2017). These relatively rapid developments have already reached a point which justifies the use of the expression ‘performance management era’ in universities. Particularly central to the performance management complex is the increasingly pervasive employment of performance measurement systems (PMS) to monitor and evaluate academic performance. Of the effects of the PMS, our focus is on the *understanding and employment of quality in academic research*.

To further specify our focus, this study is inspired by a worry that the difficult-to-define quality in research is getting increasingly and potentially too simplistically determined by its measurable proxies – for instance, based on the impact factors of journals or various kinds of journal rankings (e.g., Adler & Harzing, 2009; Gendron, 2008, 2015; Giacalone, 2009; Willmott, 2011; Humphrey & Gendron, 2015; Parker & Gurthrie, 2013). Such proxies seem to fit well with modern governance technologies and mind-sets of performance measurement in the academe, but it is far less clear whether they can capture the virtue of quality in the way that would reflect the underlying pursuit of ‘good scholarship’ (cf. Hopwood, 2007; Czarniawska, 2011; Alvesson et al., 2017).

Enhancing productivity has been viewed important at universities ever since the ideas of ‘New Public Management’ reached it (e.g., Hood, 1995; Lapsley, 1999; Modell, 2003, 2005) – and the PMS is tightly

connected to this program. The domain of performance management, and more specifically performance measurement, in the context of universities has been a target for a significant amount of academic inquiry (e.g., Modell, 2003; Sousa et al., 2010; ter Bogt & Scapens, 2012; Kallio & Kallio, 2014; Sutton & Brown, 2016; Kallio et al., 2016; Kallio et al., 2017). This literature emphasises that in attempting to increase the efficiency of university operations there is inevitably a trade-off. In terms of the production of research, it is frequently suspected that increasing the quantity of output may reduce the overall research quality (e.g., Lewis, 2014; Alvesson et al., 2017). However, while this trade-off is common knowledge, as far as we know, it has not been studied in any depth. The main issue in this regard is the tendency of using the word ‘quality’ in the academe as a taken-for-granted category, as if everybody knew immediately what it stands for. In our analysis, we problematize this view, especially with regards to the nowadays quite routinely used standard proxies in the determination of research quality.

Our analysis has at least tangential connection to the literature on incomplete performance measures (cf. Jordan & Messner, 2012). However, the incompleteness of performance measures is not key here; nobody would likely deny that all proxies are always more or less incomplete. A more fundamental issue for us is whether academics sufficiently realise the risk of the ‘tail starting to wag the dog’ and how they deal with this risk. In this context this issue means, first of all, the risk of letting the (incomplete) performance measures (the proxies) become the central target for activities rather than the scholarly research process (generation of novel ideas, development of substantive arguments, etc.) that arguably should be the focus of academic work (cf. Alvesson et al., 2017).

Empirically, we will focus much of our attention to the role of ‘manager-academics’ – heads of department, deans, and rectors – regarding the problem area. These people tend to play a central role regarding seeking, for instance, to manage the finances of the department (pressure to have a balanced budget), the volume and quality of research outputs and the educational outcomes (e.g., teaching quality), as well as care about the workplace climate. Hence, manager-academics must navigate expectations from (in principle and at least) two directions simultaneously – the administrative direction and the scholarly direction. This can easily lead to a situation with conflicting tensions.

The felt tensions have likely recently become more pressing as the degree of managerialism has increased and the intensity of the publish-or-perish culture has intensified in the academe. Not least as there likely also are different kinds of attitudes towards managerialism and publish-or-perish culture with different manager-academics, they may also have developed different kinds of strategies to handle the complex set of expectations they encounter. One of the inspiring ideas for us is the theorization of Karpik (2010) – dealing with the valuation of ‘singularities’ – which seems to be directly applicable to scholarly research. Especially the two distinctions he makes, active vs. passive and autonomous vs. heteronomous modes of evaluation are helpful for our analysis. Another source of inspiration is Brunsson’s (1985) distinction

between decision-rationality and action-rationality: we are interested whether, to what extent and how manager-academics use “short-cuts” in their decision-making and how this is implicated in their mindset and actions concerning quality. Regarding manager-academics, we will pay a lot of attention to how they sense and employ their agency in their work and how this may implicate in the practical meaning and employment of the notion of quality presuming that the actions of manager-academics play an important role in this context.

The main objective of our study is to examine the practical meaning and employment of the notion of research quality in the current circumstances of performance measurement in the academe, analysing particularly the role of manager-academics in this regard. This objective drives the following three more concrete research questions:

RQ1: How is the notion of quality understood and employed in the practice of scholarly research? What variation, if any, is there in the practical meaning and employment of quality across local contexts? To what extent is this attributable to the localised performance measurement practices?

RQ2: How do manager-academics use their agency to promote or buffer the instrumental pressures related to managerialism and the publish-or-perish culture? What effects do these efforts have on the practical meaning and employment of quality in scholarly research?

To examine these questions empirically, we will conduct a comparative case study between two largely comparable faculties from two universities in a European country, focusing on manager-academics at them, employing not only interview method but also reliance on introspection as one of the authors is currently a manager-academic, one has long prior experience from being manager-academic – and all three are academics overall (cf. Alvesson, 2003). We conducted eight interviews in each academic institution at several hierarchical levels of manager-academics, including in both cases the rector of the entire university, the dean of the faculty in focus, vice-deans for research and several heads of departments.¹

When selecting the faculties for empirical analysis, we wished to form, in principle, relatively identical counterparts allowing a meaningful comparative case study design. However, we expected that, for instance, the different trajectories of the historical development of the two faculties in question and the mere choices of personnel might have led to variation in the way manager-academics act regarding the questions of interest for us. As we will report later, while the findings offered us some notable empirical surprises, they yet allowed us to develop our main explanatory model. The main contribution of this study is to shed light on how the understanding of quality in research is affected by changes in the wider

¹ For reasons of securing anonymity, we will not use these positions in the paper as related to the quotations we will use, but call all interviewees with a general term ‘manager-academic’, and an abbreviation ‘M-A’ (e.g. M-A 1, Case A) in the connection of the interview quotes.

academic environment and local managerial practices and systems, specifically the PMS. We demonstrate how local variations in the PMS, in terms of how restrictive or inclusive they are in defining academic performance, influence the willingness and ability of manager-academics to exert their individual agency. While manager-academics were generally found to be rather reluctant to exercise their agency, or at least the bulk of them did not perceive quality in research to be a concern in which they had to actively address, a few of them from especially one of the two faculties indicated significant inclination to employ determined and long-term agentic effort in that regard. More typically, the understanding and employment of quality is lent away, to be 'automatically' taken care of through the self-governance of academics, by expectations from the discipline field, or through the local PMS – a phenomenon that forms the major worry for us when interpreting what our findings mean regarding the future of the academe.

The remainder of this study is structured as follow. In the following section we discuss changes in the global and national contexts with regards to performance measurement, the meaning of quality in research, and the agency of individual academics. Next, we detail our method, and present the empirical findings. We then discuss the implications of our results, before providing a conclusion to the study.

Performance measurement, research quality and agency at universities

The global and national performance measurement context

In the last three decades the higher education sector has been subject to significant regulatory reform. These changes have resulted in an increase in institutional complexity and raised tension between differing ideologies informing what the role of universities in society ought to be and how their activities should be coordinated and controlled (Thornton & Ocasio, 1999; Greenwood et al., 2010; Thornton et al., 2012). Reforms to the public sector, in general, and universities in particular are based on an economic rationality and exchange value of a market logic and the hierarchical control and individual accountability of a corporate logic (Neumann & Guthrie, 2002; Thornton et al., 2012; Townley, 1997). These reforms have introduced marketization of university services and the introduction of corporate sector managerial practices. This has enhanced competition between universities, departments and individual academics for students and funding, an emphasis on efficiency and productivity, centralization of authority, more active and visible control by administrators, increased application of explicit performance standards, assurance and accreditation exercises, and accountability based on measurable outputs. The felt pressures have become more persistent as the degree of managerialism and also the intensity of the “publish or perish” imperative have increased (Broadbent et al., 1999; Deem, 1998; Hood, 1995; Parker, 2011). Some have suggested that the current academic life is about competitiveness in its saddest manifestation, labelled by

skewed mind-sets and professional amnesia, and where the purpose of the academia has been forgotten (Giacalone, 2008; see also Adler & Harzing, 2009).

The pursuit of the 'world-class' university has during the last two decades rapidly spread across the globe. Excellence in research and education has become a matter of management and policy making. International rankings and accreditations have been feeding this process of positional competition (Yudkevich et al., 2016). Universities compete nowadays for the best students and faculty in order to succeed in measurements underlying rankings, other performance evaluations, and reputation building in general. As Enders (2004) notes, performance measurement in the university sector transforms qualities into quantities, making it easier to access and process information. Measurements of research quality are considered as trustworthy transparency tools not only among political decision makers and government officers, but also among 'the ordinary men', and are firmly tied to the 'New Public Management' (NPM) initiative driving effectiveness and efficiency in the public sector (Hood, 1995; Lapsley, 1999; Modell, 2003, 2005). In other words, such measurements are considered, and increasingly applied, as judgment devices of academic performance facilitating (alleged) commensurability (Sauder & Espeland, 2006; Karpik, 2010; Bialecki et al., 2017). Competition and performance measurement has potentially had positive effects on the efficiency of research productivity and thereby on knowledge accumulation. On the other hand, questions have been raised as to whether these developments have improved the quality and impact of research at all (e.g., Edwards & Roy, 2017).² A worldwide survey by van Dalen and Henkens (2012) shows that the large majority of their respondents perceive the publication pressure as high. Scholars see both the pros (upward mobility) and cons (excessive publication and uncitedness) of the so-called publish-or-perish culture.³

The national mechanisms through which public funding is allocated to universities are versatile, but it is increasingly common that they are results-based (ter Bogt & Scapens, 2012; Kristensen et al. 2011; Pettersen, 2015). The national models tend to follow the global 'corporatization' of universities and ranking

² It is also an important question how the current university environment supports the need for inter-disciplinary approaches, needed to address wider, societally important research questions. It seems quite obvious that publishing such research in the top disciplinary-focused journals may be difficult if not impossible already at the outset. The question then is: how should such research be evaluated and how it could be appreciated more and encouraged? Especially for younger scholars this is something not encouraged to pursue due to too high risks of not getting published in the best journals, which again affects the academic career negatively.

³ Concerns have also been raised on the games around how journals, especially the elite/top ones, may function and the paradigm politics around journals are well known (e.g., Macintosh, 2004; see also Eendenich & Trapp, 2018). It has also been suggested that in order to succeed in journal rankings, Editors may employ suspicious practices, for example insisting adding unnecessary as such references to the journal in question before acceptance. Concerns have also been raised related to the subjectivity of review processes (Lamont, 2009; Lamont & Guetzkow, 2016) and the fact that reviewers may request citation of their own work (Edwards & Roy, 2017). Similar signs of manipulation have been attached to university rankings, where top league universities have been noted to manipulate the information given to institutions composing the rankings (Bachrach et al., 2017).

initiatives. They are typically volume-based at the outset (mainly number of degrees and publications), and quality is induced in some of them by applying journal rankings to value the research outputs and student feedback survey results, for instance. The European country where our case study is conducted is an interesting context to examine these questions as the higher education sector has undergone significant changes in the last decade, including a notable change in regulation. The main regulatory change entailed an emphasis on the rights of individual managers at the expense of collective decision making. In other words, we may say this meant a notable shift in the university sector towards managerialism.

Rectors/Vice Chancellors/Presidents, Deans, Department Heads, and other academic managers as well as the non-administrative faculty are implicated in these developments. It seems obvious that manager-academics tend to feel a pressure to 'play the game' of global competition, the intensity of which naturally varies depending on the context and mission/strategy of the university or faculty/school. University leaders need to respond to governmental and political imperatives to gain legitimacy as the quality and relevance of research and education is constantly evaluated and debated (Watermeyer, 2014; Martin-Sardesai et al., 2017). This is important to manager-academics also because funding and resources are tied to such evaluations (ter Bogt & Scapens, 2012). While the publish-or-perish imperative typically attached to this phenomenon has raised concerns regarding quantity versus quality aspects, everything seems to yet continue the same way as the 'rules of the game' appear to be silently accepted and institutionalised (Graf et al., 2017). The leading players that are at the top of rankings do most likely not even want to go against the current regime (see Bachrach et al., 2017).

Most of the research examining the consequences of regulatory changes and the encroaching managerialism in universities has been conducted at the institutional or organizational levels of analysis (e.g., Deem, 1998, 2004; Guthrie & Neumann, 2007; MacDonald & Kam, 2007; Modell, 2003; Neumann & Guthrie, 2002; Parker & Jary, 1995; Prichard & Willmott, 1997; Willmott, 1995). This literature establishes the potential negative effects on academic life and the general opposition to increasing managerialism, but provides little insight into the effects on the micro-practices of individual academics, especially manager-academics. Survey studies in the area often generalize observations on PMS in the university sector like there were no differences in their design and especially use.

Local performance measurement systems

Performance measurement and management in universities has been widely studied in recent years. The focus points of the studies range from motivational implications of individual researchers to ideological implications at the level of the whole academe (e.g., Modell, 2003; Sousa et al., 2010; ter Bogt & Scapens, 2012; Kallio & Kallio, 2014; Sutton & Brown, 2016; Kallio et al., 2016; Kallio et al., 2017). Indeed, universities have for a long time developed measurement systems to follow and control their performance, even

before the NPM movement, although nowhere near to the same degree as today, as earlier they were typically only loosely connected to the practices. While many of these measures are related to student feedback and degree performance, the research output was also measured with quantitative instruments. Such measures are many times connected to globally applied journal lists/rankings (we refer to these as global journal ranking systems) or ones determined by national evaluation panels (we refer to these as national journal ranking systems).

While such developments may have had a positive effect on research productivity, not least as every measurement system is inevitably incomplete and tends to have some unintended consequences (Jordan & Messner, 2012), they have also introduced perverse incentives into the academe (Edwards and Roy, 2017). For example, incentives driving an increase in the number of publications may lead to massive amounts of substandard, incremental papers, reduced quality of peer review, and, ultimately, 'bad science'. Yet, PM systems remain a dominant management mechanism in universities (van Dooren and Thijs, 2010). Often, PM systems have also a direct link to funding.

Modell (2003) argues that goal-directed approaches to developing new PMS have often led to loosely coupled solutions with a view of the underlying goals. Further, Vakkuri and Meklin (2003) point out in their conceptual work that the use of PMS in universities may vary because of different cultural backgrounds. Typically universities consider them as structures directing attention rather than as accountability systems. The authors suggest that universities tend to diminish the significance of PMS by practising game rationalities and politics of representation. The former refers to 'playing games', even manipulating data, so as to have a positive impact on measures of their own performance especially in zero-sum game situations, whereas the latter denotes attempts to influence the PM system *per se*, by actively influencing the PMS criteria and objectives to be as favourable as possible from one's own perspective (e.g., regarding particularities such as publication practices of one's own discipline).

The study by ter Bogt and Scapens (2012) echoes worries regarding the performance measurement trends in universities through a case study of two universities (one in the UK, one in the Netherlands). While PMS have had a developmental role – helping individuals to improve their (future) performance – they observed that recently the systems have become more judgmental, i.e. seeking to quantitatively evaluate past performance. They claim that such PMS create uncertainty and anxiety, as they bring along risks in relation to fundamental values of science, especially creativity; how is creativity secured as researchers play safe in getting the publications they need for career purposes? One of their major arguments is that with the new PMS, the decisions (and also subjectivities) are relocated from the departments to more distant administrators.

Kallio and Kallio (2014) argue that the intrinsic motivation of academics is threatened by the new PMS at universities, especially as regards the quantitative elements of them. Their empirical findings appear yet somewhat contradictory, as for instance there are many who think that the PMS does not affect their work. The empirical findings of Kallio et al. (2017) demonstrate that universities' internal PM practices have resulted in the quantification of quality and probably also to sub-optimizing and free-riding. The former refers here to the search for 'quick results' and neglecting potentially important topics that require more effort and time to study, while the latter refers to situations where some people just want to do research – due to the incentive systems – and leave everything else for others to take care of. Free-riding can also imply that some researchers are not doing their fair share of the workload in a team, as they want to be involved in too many projects in order to publish as much as possible without the necessary time resources. The survey respondents and interviewees indeed desired for more wide-ranging PM regarding the whole, including research, teaching, and service. They also found evidence to support Lewis (2014), who argued that when quantity becomes valorized, potentially perverse outcomes emerge, including, for example, producing 'multiple publications that are small variations on a theme' (Lewis, 2014, p. 424), sometimes called 'salami-slicing' of research pieces.

Overall, the previous research provides relatively good visibility as to the landscape of performance management, and in particular PMS, in the university sector. However, we seem to know quite little about how PM systems are mobilized under 'the many faces' of managerialism, i.e. the different forms and degrees of it as applied in practice. This holds especially as comes to how one of the very basic virtues of scientific work, that of quality, is perceived and managed in this environment. We also seem to have very limited knowledge of the specific, local ways in which manager-academics apply PMS information in this regard; what is their agency like (or not) in shaping and managing their academic operating environments.

On the notion of quality

'Quality' is a term frequently heralded and celebrated in the academe. It is viewed as a good aspect of academic work overall – and not least so with regards to scholarly research, where seeking for high quality is a taken-for-granted phenomenon, if not even a constituting feature for it. It is hence unsurprising that there are, in the current performance measurement era, increasing attempts to measure, or at least control for, quality of research as part of the managerial attempts to add valuable productivity of the academic units. However, what *is* surprising, is the often silenced or omitted definitional difficulty that surrounds the notion of quality, having two very significant implications. First, any measurement attempts based on a vague definitional basis can turn out to be a futile exercise. Second, there is a genuine risk that measurements on such futile condition can start to be living their own life, leading to a situation where the

‘tail wags the (hard-to-define) dog’. Next we will take a broad look at the literature on the notion of quality, ranging from novels through business studies and higher education literatures to economics.

We start our journey from the major inspiration for our research, Robert Pirsig’s “Zen and the Motorcycle Maintenance” (1974), in which the key character, named Phaedrus, makes a very serious issue of the definition of quality, which led him, after a most profound philosophical analysis, to conclude that “true quality” is undefinable. Whether this conclusion is inevitable is not central to our analysis, but rather how Pirsig (1974) manages to demonstrate how puzzling only defining the notion of quality can be viewed. In terms of our theoretical orientation, Pirsig (1974) provides a teasing ‘way of thinking’ about quality.

To offer some more details, making a real issue of the notion of quality in the academic context leads Pirsig’s Phaedrus to first distinguish between “romantic” and “classic” ideas of quality. The former posits that quality is an immediately present and directly perceived excellence, but difficult to explicate and measure. The latter again views quality as being based on analysis, which takes potentially many relevant aspects thoughtfully into consideration, allowing them often also getting explicated and quantified. After a meticulous and mentally painful scrutiny, Phaedrus concludes that quality actually can be viewed to connect the two mentioned ideas, however at the same time still arguing that a true, “absolute quality” cannot be eventually clearly defined: For Pirsig, it is a transcendental notion that pre-exists definable and observable categories (cf. Kant’s “dinge an sich”).

Some higher education researchers have linked their analysis closely to Pirsig’s (1994) thoughts. Shields (1999) leans very much on Pirsig’s idea about the (as such not definable) notion of quality, yet somehow connecting the romantic and classic aspects. Shields notes how easily performance measurement starts leaning much on the classic aspect of quality. She however argues that the Balbridge Award⁴ is actually following the key ideas of Pirsig: Quality is about the relationship of the romantic and the classic and it is not static, but in the move. In our view, however, while there may be some validity in what Shields says about Balbridge Award, it still does not seem to be able to avoid the risk of overplaying the classic over the romantic – which feature may easily apply to just any performance measurement and development system, we suspect.

In the same area, Harvey and Green (1993) provide a thorough and many-sided conceptual analysis, essentially leading to the conclusion that quality should be viewed as a relative concept: it means different things to different people and that there is the quality of the outcome and the process. They go about examining both the philosophical underpinnings and political implications of the hard-to-define notion of

⁴ In 1987, the US Congress established the Malcolm Baldrige Quality Award (Baldrige Award). The Award's purpose was to promote Quality throughout the US; recognise outstanding organisations; and foster sharing of best practice information throughout the US. Subsequently a Baldrige Award has been established for higher education (Shields, 1999, footnote 3).

quality. Importantly, Harvey and Green note that quality is often mentioned as if we knew what it means (cf. Ball's, 1985, telling title "What the hell is quality?"). However, they also remind us importantly that being content with the philosophical position that quality cannot be defined is not so easily feasible in practice, at the same warning us about the potential pitfalls of the pragmatic approach:

"Reaching the conclusion that we might all have different understandings of quality in higher education and that none of us is necessarily wrong or right does not mean, however, that we are absolved of the responsibility for maintaining and enhancing quality. It is merely the adoption of a pragmatic attitude. In practical terms decisions have to be taken: courses have to be approved or turned down, funding has to be allocated, new lecturers have to be appointed in competition with others. The pragmatic approach determines a set of criteria that reflect common-sense aspects of quality and then seeks out convenient measures by which to quantify quality. Unfortunately, this approach sometimes works in reverse. Convenient measures are eagerly seized upon and a rationale constructed to give them credibility in measuring quality. The *ad hoc* use of performance indicators is a case in point." (p.29)

This is not least since, Harvey and Green argue, the "value for money" take of quality in higher education became stressed by the political right in the UK, starting in the 1980s. Thereby seeking efficiency, effectiveness and accountability became increasingly important in the public sector administration, including universities. In the name of the 'New Public Management', similar tendencies have emerged also thereafter in other EU countries and in the US.⁵

The relatively recent piece by Wittek and Kvernbekk (2011) is especially relevant and helpful. A quote from their abstract is informative regarding their main arguments:

More specifically we discuss the possibility of obtaining a precise, unified definition of quality by addressing the problem of asking "what is" questions. We use definition theory and theory pertaining to linguistic vagueness, including Wittgenstein and the idea of family resemblance. We suggest that quality is an inherently vague concept, which runs us into boundary problems and forces us to operate in grey zones. This means that if your "what is" question is a question for the essence of quality, you make demands that the concept cannot fulfil. Recognition of this fact

⁵ There is a wide set of pieces that adopt an even more pragmatic take on quality, implicitly mostly reproducing the classic alternative (in contrast to the romantic one) of quality of Pirsig (1974), often leading to listing of aspects of quality. Garvin (1984) for instance suggests five approaches to quality (of a product) and argues that different functions of a firm tend to stress different approaches and dimensions of quality. A firm can try to produce high quality products regarding most/all of these dimensions (e.g., Steinway pianos) or just a few of them, like reliability and conformance (e.g., Yamaha pianos). Another rather similar example is Reeves and Bednar (1994), who suggest that quality can be seen alternatively as excellence, value, conformance to specifications, or meeting and/or exceeding expectations.

implies that our expectations concerning concept precision and unity should be modified. (p. 671)

Accordingly, Wittek and Kvernbekk argue that quality cannot be clearly defined, because the notion is ambiguous (as it has many meanings) and vague (it is difficult to decipher at its boundaries). Situational stipulations help in the former regard, but regarding the latter, we just have to accept the uncertainty around the notion. Wittek and Kvernbekk borrow from Wittgenstein (1953) the idea of ‘family resemblance’ suggesting that there is still something that the various stipulative expressions of quality refer to, even if having potentially multiple meanings and being vague at their very boundaries.

The spirit of Wittek and Kvernbekk (2011) that we also readily share comes well afore in the following quote:

[...] it seems to us that even in the absence of an agreed-upon, unified definition of quality, we all (think we) recognize quality when we see it. Harvey and Green [1993] seem to connect recognition with the first of their proposed categories, but a case could be made that this holds for all domains where the concept is used. We can tell the difference between good and poor student papers when we see them, even if we cannot pinpoint exactly the basis of our judgment. Art experts agree that one painting is better than another, even if they can point to no objective criteria. This is interesting, given the lack of a clear definition of quality. We still (think we) know what it is. (p. 675)

We feel the analysis of Wittek and Kvernbekk (in the spirit of Pirsig) offers a very helpful basis for our exploration as it stresses that quality cannot be clearly defined and it might be wisest to just accept this. While their piece does not progress much over and above just ‘proving’ this, it still offers us – jointly with Pirsig (1974) – a good platform for our empirical analysis. We will work along the arguments of Wittek and Kvernbekk (2011) that while quality is arguably both an ambiguous and vague notion, and we have to live with that, referring to Wittgenstein’s (1953) ‘family resemblance’, there still is some common idea keeping the various formulations of the idea together. For our analysis in the context of scholarly research, we will take the Pirsigian not-strictly-definable-excellence as the core of the family-resembled variations of quality.⁶

When turning to economics, Karpik’s (2010) recent book “Valuing the unique. The economics of singularities” presents a notable new opening and problematization to this very problem area. It is worth noting that Karpik mentions that he was contemplating between using ‘quality’ or ‘singularities’ as his

⁶ When outlining the scope of ‘good’ motives (other than just ‘playing the game’) of researchers, Alvesson et al. (2017) list the following features: “intellectual curiosity, open-mindedness, willingness to reflect and challenge one’s own and microtribes assumptions and vocabulary, to take risks, to make ‘non-convenient’ empirical work, to devote much time to interpretation and writing, and possibly postpone finishing the PhD or a promotion until one has something important to say” (p.66). Our analysis in this study draws on similar kinds of ideas regarding what quality in research could encompass.

anchor term, finally choosing the latter, but inferring that the same analysis would apply had he used the term 'quality'. The reason for choosing 'singularities' ('the unique') is relevant for our analysis:

"For a long time, I used the term *quality* (or *qualities*) and its derivatives *quality goods* and *economics of quality* or *qualities* [footnote]. But the word's affinity with a unidimensional reality, its increasingly frequent use, the growing diversity of its meanings and the misunderstandings it prompted led me to replace it with the notion of *singular products* (goods and services) or, more simply, *singularities* [...]" (p.10)

For Karpik, singularities are goods and services with three main features: (1) they are multidimensional, in that they are made up of various overlapping dimensions; (2) they are incommensurable because there is no common and obvious measure that enables an indisputable evaluation of their quality; (3) they are characterized by uncertainty, as evaluating quality of singularities is complex and often cannot occur until the good is bought or the service is provided. The main motivation for Karpik to very thoroughly examine singularities is that, he argues, neoclassical economics keeps silent of them, therefore omitting an important part of economic life as an uncharted territory. Karpik's analysis of singularities is relevant for us since only with slight adaptation we can transform these thoughts, explicitly dealing with singular products and services, to a piece of research as the 'good' in focus.

Importantly, Karpik suggests two analytical dimensions for a quality assessor's action (types of involvement) for being relevant therein: active vs. passive and autonomous vs. heteronomous judgment. Using Karpik's theory, we could say that we are especially worried of passive and heteronomous take on assessing research quality. This refers, broadly speaking, to "outsourcing" one's own judgmental work to some outside party (e.g., just take global journal rankings, many times drawing mainly on impact factor measures, and assess published research based on that), while we like to support active and autonomous judgment: the evaluative interest focuses directly on the substantive contents of the piece, and while perhaps using some rankings as wide reference, essentially making of an independent overall judgment. Karpik also carefully analyses the various kinds of techniques and technologies – which he calls "judgment devices" – offering short-cuts for assessors. Journal and university rankings are representative examples of such devices.

On human agency

The question of agency is closely linked to the rather philosophical question of free-will of humans, which has so far been unresolved, but has been near the heart of the so-called Science Wars (e.g., Gross & Levitt, 1994), for instance. Those who believe strictly in determinism argue that there are certain mechanisms (founded on physics, chemistry, biology, psychology, sociology etc.) that govern our behaviour and it is only a question of time, and advances in science, that we will understand all of it. Taking this approach quite

seriously would mean that causality is always just deterministic and there is no true room for free-will. Towards the other end of the continuum, there are those who believe that while many such mechanisms can perhaps be discovered in sciences concerning non-human objects, human action will always remain at least somewhat discretionary, leaving room for human free-will and individual agency. The latter view leaves space for a difference between determinism and causality – and for the view that causality may exist jointly with people’s agency (Melan, 2013; Lukka, 2014).

Leaving this level of dispute regarding human free will aside, in many approaches to social sciences it is typical to lean, more or less, towards the latter view, treating humans at least *as if* they had agentic discretion in their action. One of the key social scientists regarding agency is obviously Anthony Giddens. In his structuration theory (e.g., Giddens, 1984), knowledgeable agents are often able to conceptualize and reflect upon the practices they entertain, leading to the feature of double hermeneutics. Accordingly, agents can, for instance, trigger or hinder the actualization of particular, presumed causal mechanisms to implement their agenda. However, they probably cannot, at least in the short term, change the causal mechanisms themselves (Durand & Vaara, 2009). That said, since in line with the defining feature of social constructionism, agents often have the possibility of acting otherwise (e.g., Hacking, 1999), knowledgeable agents’ intentions and actions are potentially changeable (Lukka, 2014).⁷

In our research, we are interested in the perceived agency of researchers – especially that of manager-academics. What is the role of their agency in managing and leading their organisations and their subordinates? And in particular, what is their role in leading their organisation’s members’ thoughts, understandings and actions regarding quality in their research? Quite recently, Alvesson et al. (2017) have paid attention to the issue of academics potentially not being aware, or not using to its full potential, the agency that they actually can well have at their disposal: “Many academics downplay the freedom they have in constructing their identity” (ibid, p. 100) and “most researchers have more power than they realize” (ibid, p. 139) as it is precisely academics themselves who evaluate research and researchers as well as take part in academic management, funding bodies and advisory boards. Our empirical study will shed light on these types of matters, in particular, how manager-academics either reproduce, or challenge, those structures (i.e., PMS) that lead to research quality being understood and employed in particular ways.

⁷ Agency is very close to the notion of human reflexivity. Some social scientists have criticized Giddens being underdeveloped regarding agency/reflexivity (Archer, 2007; for a reconciling position, Stones, 2005; see also Modell, 2017).

Summary

Our working model for the study indicates that there are four main concepts that are of interest: 1) Global and national performance measurement context, 2) Local performance measurement systems, 3) Manager-academics' agency, and 4) The practical meaning and employment of quality (Fig. 1). The model implies two empirical layers. At the general level, there is the global and national context of increasing managerialism and publish-or-perish culture, and the general effects this has on the understanding of quality. Much of this relates to the increasing demand to quantitatively measure quality to make comparisons and rankings between universities, schools and individual academics. At the more specific and local level there are different universities and schools, and the variations in structures and practices, and the degree and use of agency by manager-academics. The arrows of Fig. 1 represent the influences and their directions which were derived by integrating the literature review and our own empirical findings (cf. the discussions section below).

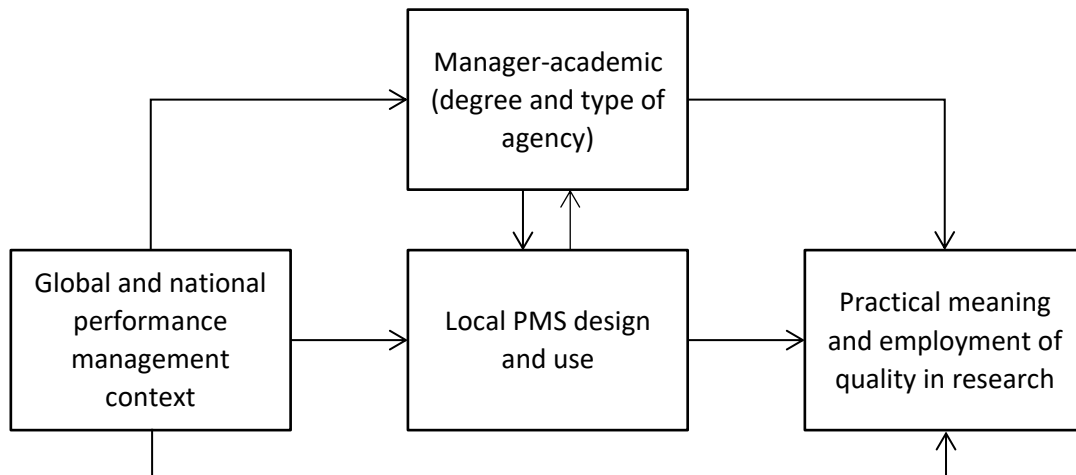


Fig 1. The research setting.

Methods, research materials and analysis

Our study explores the practical meaning and employment of the notion of research quality in the current circumstances of performance measurement in the academe, focusing particularly on the role of manager-academics in this regard. While the starting points for our study were some broad deductively tuned discussions among the research team regarding the interesting themes in the area of performance measurement – all of us had earlier conducted research in that area – overall, the research was characterised, even dominated, by abductive features (Dubois & Gadde, 2003; Lukka & Modell, 2010). Accordingly, once we had decided to examine performance measurement at universities and then started

our interviews, then the research process was very much of the type typical of abduction: swapping continuously between the emic and the etic domains towards our conclusions at the etic level (Jönsson & Lukka, 2006; Lukka & Modell, 2010). The overall spirit of our study is interpretive – capturing the meanings of the people interviewed plays a notable role in this study. Despite the fact that the research process was generally abductive, we chose to write the paper up in a deductively reconstructed manner to improve its readability.

The research is empirically designed as a comparative case study, where the target organisations were two university faculties in a European country. These two faculties were selected, firstly, due to the fact that they represent the same major disciplinary area with quite similar focus and scope and, secondly, due to the expectation that there would be differences in the ways how manager-academics are acting at them due to several reasons, which were initially naturally not much more than *ad hoc* inferred guesses. We interviewed altogether 16 manager-academics, eight from both organisations. The interviewed people included rectors of the respective universities, the deans, the vice-deans responsible for research and many heads of department. The two sets of interviews were forming a good match with a view of the positions of the interviewees.⁸

We conducted the interviews using a carefully prepared semi-structured research guide (interview guide in Appendix 1), which we used, in line with the idea of theme interviews systematically, still allowing room for free association by the interviewees as well as for asking further questions on topics that happened to surface. At the start of the interview, we informed the interviewees only about our general research interest, to control for any potential ‘halo effects’ as well as, importantly, promised them anonymity. The interview lengths ranged from 70 to 132 minutes, the average being 96 minutes. 14 of the interviews were conducted face-to-face and two through Skype (more detailed information on the interviews in Appendix 2). The interviews were conducted in most cases by two of the researchers of the team, a few times by all three and a few times by one of us only. All interviews were recorded and transcribed verbatim.

As for the analysis of the interview materials, a brief note was made after each interview to our research diary to capture immediate impressions and observations. Then, once the interviews were transcribed, we prepared memos (of two to four pages) on each of them, which focused on the perceived highlights of each interview and consisting some of the most helpful or exciting quotes. Our analysis was generally triggered by our shared broad worry concerning the destiny of research quality under the current condition in the academe. In line with this, essential for our analysis was to make and keep it consistently clear what is the intended *explanandum* (cf. Lukka, 2014): how research quality was given a practical meaning and how that

⁸ For reasons of securing the promised anonymity for all of our interviewees, we are in the write-up of this study especially careful concerning what we disclose about our interviews and interviewees.

was employed by the manager-academics we interviewed? In a close dialogue of going back and forth between our original theoretical ideas, concepts, data, and interpretation (Ahrens & Chapman, 2006; Lukka & Modell, 2010), we gradually developed our explanatory model for what drives the ‘practicalisation’ of quality in research as depicted in Figure 1.

The entire research process was significantly facilitated by a careful keeping of a research diary, consisting of chronologically organized notes on everything that we thought mattered in our research: Many kinds of notes on our numerous brain-storming sessions concerning our main theoretical ideas, the research question and its motivation; immediate impressions right after each interview; and notes on the development of our theoretical storyline and conclusions over the entire abductive process. The theoretical storyline of our paper advanced and matured already in the research diary, which greatly facilitated the write-up stage of the paper.

The two case studies

Performance measurement and management systems of the two cases

In both cases, simply termed Case A and Case B, signs of increased managerialism were observable, as well as a tendency to acknowledge the publish-or-perish culture being the ‘new normal’, with interviewees noting both positive and negative effects. Typically, the good thing about this was seen to be making it clear that every faculty member needs to do research/publish, but also that everybody is expected to aim as high as possible regarding publication outlets. However, the downside was also generally recognized and summarized to relate to the potential temptation to use short-cuts in making science under notable time pressure. Demonstrating well the general idea, one of the interviewees at Case B reflected on personal experience regarding an ‘unmatured’ piece of work that was submitted to a journal no matter of the worries raised by several colleagues:

I was sort of in the opinion that if we are striving for quality, those [GJRS1]⁹ publications don’t come in two years, they normally take more time and so actually you should allow people to build up a strong empirical base and then work on that base and then publish and see what the results are, instead of saying that ok, you have to get something done in two years’ time. Like our Dean seems to be in the opinion [...] I can see an example from my department right now when [the academic’s] tenure came up, they rushed to send some stuff to a journal. Which I wouldn’t have advised them to do so but then again what can I do? (M-A 4, Case B)

⁹ One particular global journal ranking system; labelled here as GJRS1. In order for a journal to get listed in GJRS1 it needs to reach high values in bibliometric analyses, but there is also a voting mechanism where selected university faculties have a vote. The journals included are geographically highly concentrated as comes to their editorial teams and publishers. The list also reflects paradigmatic homogeneity to a notable degree.

In both of the analysed cases academic performance is carefully followed, formally and informally. In Case A, a PMS, trying to capture both the quantity and quality of academic activities, has been in place for almost two decades. The idea of the system has been to signal expectations and making visible the multi-dimensionality of academic work, as well as respect for such versatility. The system allocates points across a wide range of academic activities, such as journal editorships, journal reviews, conference organisation, external funding, citations, as well as publications. In Case B, the design of the formal measurement of research performance is based on highly ranked publications. Both cases have thus introduced measures thought to capture research quality, yet with different emphases. In Case A the PMS is tied to a national journal ranking system (NJRS).¹⁰ In practice, the system aims to signal also that publications at lower levels of the ranking than the highest level are also highly valued. In Case B the main emphasis is on global journal ranking systems (GJRS) and a crucially important part of the PMS configuration is the tenure track system, in which one particular GJRS, GJRS1, plays a very important, even decisive role. Table 1 below summarizes some of the core aspects of the PMS of the two organizations regarding research performance.

¹⁰ The basic principle underlying the NJRS is to support the assessment of research output quality (in addition to measuring it quantitatively with bibliometric analyses) also qualitatively across all disciplinary fields in order to account for the different publication cultures and modes of appreciation of research quality. The evaluation work is carried out in disciplinary expert panels, to which all academics can submit well-reasoned suggestions regarding journal classifications.

Table 1. The basic characteristics of the research related PMS in the two case organizations.

	Case A	Case B
Scope	Research points from publications, professional academic tasks, WoS citations, and external funding. Publication numbers vis-à-vis the NJRS classification are followed separately as well.	Primarily publications, GJRS1 publications emphasized. Other academic tasks/performance and citations are discussed informally (e.g., in salary negotiations)
Publication categories used	NJRS (used also to score academic tasks such as Editorships, e.g., being the Editor of a top level NJRS journal gives more points than in case of a lower level journal)	GJRS1, GJRS2 ¹¹ , NJRS
Connection to basic (state) funding	Strong link because the state funding model is also based on the NJRS	Indirect, but as the NJRS strongly correlates with GJRSs, there is a link
Career paths/development	Multiple paths, wider consideration of the publication (and other) record	Tenure track the foundation of the whole HRM system, GJRS1 publications very important/decisive within that in general
Other rewarding	Monetary reward for highest level NJRS publications (before certain budget cuts, used to be also for the second highest level publications)	Monetary reward for publications in two GJRSs (for GJRS1 eight times more, and for the highest level in another GJRS four times more, than the reward received in Case A)
Time orientation	Figures calculated annually, in relation to budget funding 3 year averages applied	Figures calculated annually, in relation to budget funding 3 year averages applied

Regarding connections to the results-based state funding model, both organizations have designed the internal fund allocation model quite similarly. In Case A, the main part of funding goes to Departments and other units according to a history-based resource allocation/need scheme. Just less than 10% of total funding is allocated based on the unit-based results, where the research points count for 50%. In the publication scoring, a top level NJRS publication is awarded a number of points that is one third more than what is awarded to publications in the second highest level and four times more than what is awarded to the third highest level. Other academic activities score for less, but are also connected to the NJRS, where appropriate. For example, being the Editor-in-Chief of a journal ranked at the highest level gives annually a number of points that is 50% more than the score for the second highest level and three times more than

¹¹ Another globally widely applied journal ranking system, labelled here as GJRS2.

the score for the third highest level. Similar kinds of principles apply to Editorial Board Memberships and Review tasks. Each Web of Science citation gives points in the system, and their share of the total points can, in a few cases, be substantial. This is yet considered as very important, as citations are seen as the only available proxy for research impact. Overall, the system makes visible academic activities widely, and one can occasionally score high even without publications. This is why the total score is always analysed by point categories, as well.

In Case B, the main part of basic funding is tied to the number of tenure track positions in each department or unit. The results-based share of total funding is around 15%. Out of that about one third is based on research output, where the refereed publications are given multiples. The highest multiple is attached to “Top” (as officially labelled in the model) level journals (GJRS1/highest level of the NJRS/highest levels in GJRS2). “Medium” (official label) journal articles (second highest level in the NJRS/second highest level in GJRS2) have a multiple half of that of the top level, and “Normal” (official label) journal articles (third highest level in the NJRS/third or fourth highest level in the GJRS2) a multiple one sixth of the highest level. Here we can see a difference between the two cases: in Case A there are not so big relative differences between valuing ranking categories as there are in Case B. In addition, the research PMS of Case B only includes publications, whereas in Case A the system also captures many other research related outputs and activities.

The interviewees’ perceptions varied quite a bit as comes to how useful they perceived PMSs to be, both within and between the case organizations. One interviewee at Case A saw the PMS as an integral mechanism to signal that academic performance beyond journal publications is valued. Furthermore, the system is not primarily designed for formal performance evaluations, but as a tool for self-reflection. Nevertheless, journal rankings are an important component of the system:

I think that has been instrumental or somehow a tool, a signaling device for paying attention to both quality and quantity. And of course over the years there might have sometimes been some misunderstandings that it's more for quantity and how I get more and more points. Because [...] if you really wanted to maximize the number of points, you could really do it at the cost of quality. But now the recent discussions and the recent revisions to our system really emphasize more and more [research quality], for instance, you get relatively more points now for [the highest level of the NJRS] and [the second highest level of the NJRS] publications which is a clear signal in my view. (M-A 2, Case A)

No matter of the purpose of the PMS serving as a multi-dimensional signalling device, some of the interviewees at Case A still perceived it being more like a ‘results-table’ without direct or systematic effect on management and behaviour. The majority of the interviewees at Case A though noted it to be a valuable tool for annual development discussions, for example, in addition to other information (e.g., course feedback in education). The following excerpt from the interview of M-A 7 (Case A) well demonstrates the

perceptions more generally as it brings out both the positive things and the fact that in every-day practice the use may be limited:

Well honestly it doesn't play that big role at our department. [...] these points and scores are computed every year and then... we can discuss, you know, our research output in a way that we would do anyway. So we of course pay attention to what people have published and that's the, that's where our focus is. But the other part, I mean all this, well the points count from so many sources that [...] it is true that sometimes we discuss them but we don't, I don't use these points systematically [...] anyway this information is collected and it may be informative sometimes but not these points as such. [...] I don't see that it's causing any problem or trouble. I like the idea that, you know, sort of, detailed collection, data collection is conducted, done every year. So it's a nice thing because anyway it's, when you collect those, this information it has an impact on people. They know that these type of things are measured. So they know that it counts, it matters, if I publish it matters, somebody cares. So it is important, I mean, it's a good thing. (M-A 7, Case A)

In case B, there were several positive remarks on the ways in which the PMS was employed in general. While the interviewees generally accepted the aim for 'world class' to be an important thing to pursue, and understood how the imperative for GJRS1 publications aims to drive this as part of reputation building, the behavioural implications of the pervasiveness of this initiative raised concerns, even quite heavy critiques. One of the most critical and cynical interviewees (M-A 8, Case B) described the situation in terms of 'unnecessary bureaucracy' in the following way:

Interviewee (IE): This idea that I am a goal setting manager who's going to tell you what to do next year, it's ridiculous.

Interviewer (IR): What about in terms of what you are held accountable for from the dean. How are you evaluated?

IE: Well he's pretending to evaluate me and I'm pretending to be evaluated.

IR: What is he pretending to evaluate you on?

IE: He is looking at the numbers of the department [...] one number went up or down and he said that's good or that's bad and I say actually that's not good or bad, that is just random chance that anybody with a basic understanding of statistics will understand and he said ok. That's about it.

IR: How frequently do you discuss or do you interact with the dean?

IE: As rarely as possible. [...] Well, it's a game because I know what he's going to say. I know it doesn't really matter. I know what I'm supposed to say [...] it's just for show. It's like airport security check, they have security theatre, and we have performance theatre.

This interviewee continued by explaining that the administrative performance expectations were seen as 'destructive' to the culture that had been fostered, which privileges academic freedom and conducting quality research that has impact – both scholarly and in practice. They also noted that what truly matters to academics is how they are perceived by others in their field:

IR: You mentioned this corporate culture in the administrative side. Has that had any effect on the research culture here in the department?

IE: No because we try to keep them away. We see most of what administration does as destructive and irrelevant to what universities are all about so our role is to do our own stuff as well as we can — research and teaching — and try to keep the administration away.

IR: How would you describe the culture here in the department?

IE: It's ambitious, we want impact [...] we don't care about numbers, we don't care about quantity. We care about doing a little bit of something that matters.

IR: And you see this as being in line what [Case B] wants to achieve?

IE: Partially but not much. Our University pushes much more towards quantity, also in terms of doctoral education [...] This is about one of these examples of double-speak, that the university and the [faculty] say that, of course we don't really think this, but you are supposed to get the numbers because that's where the money is from. So they are quite open about it, they realize that it's kind of a lie that we have to pretend to produce so many PhDs and so many articles that nobody reads. They think we should be doing all these things that matter but that's not what pays the bills. [...] To advance in your academic career you are much more dependent on what people in your field think of you abroad than you are of what some administrators think here in [this country], that's very rarely relevant. So the culture of your academic field always is the stronger one — that's what rewards the quality in the end.

Finally, we may note that especially the manager-academics with highest organizational ranks admitted to feel considerable pressure to produce in addition to high quality also high volumes. The same applies to the concern for impact, also other than that measured by citations (so called practical impact vis-à-vis the society). This is understandable though, as the funding models and mechanisms emphasize volume-based metrics. Impact, again, another hard-to-define-and-measure concept, is being evaluated and required by most external funders. It seems that university leaders thus need to pay attention to volume and impact, too, as it affects their resources and legitimacy in the eyes of certain stakeholders.

Findings on the practical meaning and employment of research quality

Manager-academics at the two cases expressed reasonably consistent interpretations of the notion of research quality. In general, the notion of quality was perceived as a fundamental aspiration for researchers, with some manager-academics even considering that “quality is everything” (M-A 8, Case A).

Two main aspects of quality were expressed. One aspect, noted by nearly all interviewees, was that for a piece of research to have quality it must conform to fundamental research standards and principals (i.e., “quality as conformance”). A department head from Case B referred to this as “basic quality” which comprised “must-haves”:

I think one aspect of quality are those must-haves. I mean, what is quality basically? Some say quality is when your expectations are met or exceeded and I think in terms of expectations or must-haves, it's solid good work, again, craftsmanship in terms of rigor. So conceptual rigor, methodological rigor, that concepts are properly defined or that the research question is properly defined, the concepts, the data collection or the empirical setting is properly described and data is reported. So I think this is... I get angry when I have the feeling that, let's say, this basic quality is not achieved, because this is what we should be able to produce as good researchers. (M-A 5, Case B)

A department head from Case A referred to this kind of quality as “engineering quality” (M-A 5, Case A) in which the research conforms to particular specifications that are expected at a minimum for any type of

research, such as ensuring the credibility of the data, communication of method choices, and adhering to ethical standards. Another interviewee referred to “hygiene factors” such as “good dataset, good analysis, good writing, good flow” that should be present in any piece of research (M-A 2, Case A). Other interviewees from Case B emphasised that the research should follow the appropriate “scientific method” (M-A 3, Case B) and that there is a strong methodological foundation so that “you can trust the results” (M-A 4, Case B).

The second main aspect of quality relates to the overall contribution of the piece of research. As one interviewee put it succinctly, “all that matters in the end is the contribution” (M-A 2, Case A). A contribution can, however, take many forms. The following reply, when asked about the meaning of research quality, from a department head of Case B captures many of the components of contribution commonly noted by interviewees across the two Cases:

I think, I'm struggling not to say 'contribution' because that's the password that people are using all the times but maybe that should come as the first one. So what's a contribution? So I think there's an element of newness, element of relevance, we learn something that we didn't know. Often times this means this can be counterintuitive. (M-A 6, Case B)

Most manager-academics we interviewed commented that research of quality must contain something that is really new (i.e., “quality as excellence”). One referred to this type of quality as “superior quality” (M-A 5, Case A), while another termed it the “wow-aspects” (M-A 5, Case B). This might include novel theory that opens up a new stream of research, counterintuitive or surprising findings, exploiting a unique dataset, developing a new concept, problematising the status-quo, or in general, making you think differently on a topic. While it is difficult for any one piece of research to deliver on all of these facets, “one or two of those” (M-A 5, Case B) are required for it to be of high quality.

One difference between these two aspects of quality is the difficulty in assessing whether they are present or not. A interviewee from Case A referred to quality as conformance as the “narrower take on quality” (M-A 4, Case A) that can be considered a check-list of items that should be ticked off for the research to meet an acceptable standard. This aspect of quality was viewed as more explicit and objective. One interviewee commented that researchers can be trained to conduct research that meets this type of quality, whereas in contrast generating novel ideas is something that “you cannot train” (M-A 5, Case B). Interviewees found it difficult to explicate how they demarcated this form of quality, with a number indicating that it was not something definable but more a “feeling” (M-A 4, Case A) that they had about the piece of research at stake – the research is exciting, inspirational, and evokes “wow-feelings” from the reader.

[...] an exceptional piece is something that you still are... you almost can't get sleep in the evening, it was so... there is something so striking, so new, so inspiring you just know it when you see it. So maybe the degree of novelty or the originality of the contribution, but sometimes it's quite difficult to draw the line. (M-A 2, Case A)

When asked what quality of research means, one manager-academic in Case B responded that there is “no way of defining it”, but “you know it when you see it” (M-A 7, Case B). This alludes to quality, at least in

terms of this second aspect (i.e., “quality as excellence”), as being inherently based on tacit understandings.¹² In fact, a number of interviewees pointed to the notion of quality as a socially constructed understanding within a community, and that you could only understand the meaning of quality if you are embedded within that community. This points to the difficulty of assessing quality by an outsider, with little tacit understanding of what constitutes quality within that community.

[...] the idea of quality should come from the field and from the research community and you should be able to understand what the quality means in that respect as a researcher. (M-A 6, Case A)

There's an interesting book by Michèle Lamont within science and technology studies on how, in evaluation processes, how do those who do evaluations can make the distinction between good and excellent [...] [it] tells about how difficult it is really to pinpoint the particular issues that you have to have present in making the distinction between good and excellent... it relies on your own expertise within the field. That's very key there, and that's why you need, for example in evaluations, you need to have panels [...] to be able to justifiably argue for excellence (M-A 8, Case A)

One interesting observation from an interviewee at Case B was that even though the topics that are considered important or relevant change over time, the intrinsic aspects of quality are invariant.

So whatever was high quality theoretical research in 90s is still going to be considered high quality but is this [the research question of the study] a relevant question right now, that could have changed. Maybe enough things were answered in between or maybe the world changed a little bit. What was interesting then is not interesting anymore. I think those sorts of judgment of interesting question is more volatile than the sort of more fundamental values, quality. (M-A 7, Case B)

Manager-academics at both institutions commented that high quality research was rarely of the type that generates just incremental knowledge within an established field. Some noted that such research had its merits (it is “good research”, but not of high quality), while others expressed concern about the increasing production of research that had very high quality in terms of technical execution, but had only marginal contribution to the knowledge of the discipline.

It's not just kind of testing a hypothesis in a bigger model or it's not just adding bits and pieces to existing knowledge but it's somehow questioning are we doing the right things. (M-A 3, Case A)

However, some interviewees appreciated also smaller contributions, viewing them as building blocks potentially needed for bigger contributions in the future.

Good research is something where you take an existing framework and work really hard on the foundations of it and see that ok, this is what we had to assume before, now I can do the same, with slightly weakened assumptions, for example. So this is the, you know, shovel work that just takes the frontier of knowledge further away. Little by little. It's-... I'm very happy myself also chipping away little chunks. You know, every now and then doing that you see something big and breaking away, but I'm not saying that sort of things that did not result in major breakthroughs are somehow worthless because, no, it's all sort of very small accumulating process of knowledge. (M-A 7, Case B)

While manager-academics at both Cases mentioned relevance and impact of the research as a marker of quality, interviewees at Case B tended to put somewhat greater weight on the need for the research to

¹² Interestingly, manager-academic 7 at Case B referred spontaneously to Pirsig's book (1974) when s/he started commenting the question as for what 'quality' means to him/her.

have practical or societal value. Manager-academic 2 of Case B was particularly critical, commenting that “top-notch research [is] not very important, unless you influence society through your research” (M-A 2, Case B). Another manager-academic at Case B commented:

I also sort of think that the research to be good should also speak to practice and I'm not saying that it has to be practically relevant directly, but topics and ideas should be somehow related to the idea that we could make use of that [...] So again I think when we do research that it should be somehow useful for either those who decide on laws and regulations or those who manage companies and so forth. (M-A 4, Case B)

Yet another interviewee from Case B replied that they try to always “come from an interesting, practical, real-life problem” so that their research “has meaningful managerial implications” (M-A 5, Case B).

A second point of difference between the two cases related to how manager-academics evaluated research quality. In Case A, manager-academics repeatedly mentioned that there is, and should be, a judgmental element in determining the quality of research. A interviewee at Case A noted bluntly that “quality is not [NJRS] classification” (M-A 7, Case A), one noted that quality was not determined “by the journal where it is published” (M-A 2, Case A), while another lamented the increasing reliance by the research community on journal rankings for determining the quality of research:

As a community, we have simplified it a lot. That's okay, if it's published in a good journal, it must be good. But that's not necessarily the good thing. Because not everything what is kind of in good journals is very great, and not every great piece of science is published in the best journals. (M-A 5, Case A)

The top manager-academics at both Case A and Case B commented that journal rankings served a useful purpose in gauging research quality. Metrics such as journal rankings inevitably have to be, at least to some extent, relied upon because manager-academics that sit at a distance typically have less discipline specific knowledge.

I think many things would actually need to be evaluated and addressed at the level where you have the level of competence. The level of competence is negatively correlated to where you are in the organization. The further away you are from that area, the less you know about it. (M-A 2, Case B)

The corresponding top manager-academic at Case A commented a number of times that “objective measures” such as journal rankings should be seen as just a proxy, and not equivalent to, the notion of quality. This interviewee was also of the view that such proxies are needed to legitimate his/her unit’s performance towards those who are even just a bit more distant from the unit (like the top management of the university). Manager-academic 2 at Case B, while realizing the importance of discipline specific knowledge, commented that when they are required to make an evaluation, and in the absence of other information, they rely on publication numbers and outlets to assess the capability of a researcher to produce quality output. This is particularly important at Case B because high performance is defined “in terms of GJRS1 publications” (M-A 7, Case B). There appeared to be, in some cases, contradictions between the intrinsic values and beliefs of the manager-academics in Case B and their actual practice — one interviewee held the following view of journal rankings:

I think this is good for communicating some kind of intent that we really take seriously the idea that you should publish in top outlets. It should never be used as a substitute for actually sitting down and reading the papers and making a judgment on the quality of the people that you are evaluating. (M-A 7, Case B)

Yet when asked about how they evaluate the performance of staff in their unit:

So [sigh]... I have to rely on this performance in terms of the journals in which you get it published as a signal of that because I can't judge every field of [the discipline] myself as that being interesting or not so interesting. (M-A 7, Case B)

Findings on manager-academics' agency

Limited willingness to take a position as a manager-academic

A common theme that emerged from the manager-academics at Case B was that managerial positions were taken reluctantly.

To be clear I'm not intending to continue in this role either for a long time. So I tried to decline but then we had a compromise which is to do this for one year. (M-A 6, Case B)

I was not applying for this job. I was dragged into this job. Not dragged, but I mean, on the third effort I agreed to an interview and ok, I took the job. (M-A 2, Case B)

The department head is not something that everybody wants to be. It is not a sought-after position [...] I'm sort of hoping that at the end of this year [another professor] will come back to the department and [they] will take this so I can get rid of this duty. So, again, I don't kind of feel that these are the dream jobs to be the department head in academic community because I think most of us are self-driven in the sense that they don't need much of a leadership. (M-A 3, Case B)

Interviewees in Case B primarily saw their identity as that of being a researcher – taking a managerial position is generally an unwanted obligation. In Case B most of the departments have a rotational system such that the responsibility is shared amongst senior academics over time:

I'm looking forward to sort of returning back to just basically doing more of my own research in a few years' time but everybody has his or her turn being the head of department [...] this is something we pretty much operate here on seniority. So people ahead of me in seniority have already done their deed. So it was my turn now. (M-A 7, Case B)

Taking a managerial position meant making trade-offs between what they should or could be doing as a manager-academic, and what they actually did, in order to preserve time for research related activities. One department head at Case B commented that "I'll play the game, I'll try to do my best here so maybe thirty to forty percent of my time" is spent in the managerial role (M-A 6, Case B). But they noted that they could easily spend their entire time attending to managerial responsibilities as there is "an endless list of things to do if one really wants to be a good academic leader so I totally think that, I could list a lot of things that I should be doing but I'm not, I'm only doing as much as I can" (M-A 6, Case B). Not surprisingly, then, the main tension between managerial responsibilities and research activities was time.

The biggest tension for me is all this administration seems to be eating up a lot of time and that is away from why I joined the academy at the outset. I'd be rather doing something else, that something else being research. (M-A 3, Case B)

In all honesty I think that an active research/teacher can never be a good head of department [...] there just isn't enough time to do things as well as they should be done in terms of managing people, helping them, thinking about the future, planning, creating new ideas, projects, developing the organization. (M-A 6, Case B)

The same tension between academic endeavours and managerial responsibilities was evident in responses from manager-academics in Case A.

You have to sort of accept the fact that you can't do both perfectly [...] So you just need to make compromises and I actually have even said to my people that I admit, acknowledge that, I can't be a perfect head because then I would need to sacrifice my research completely, I mean too much. (M-A 6, Case A)

However, there was greater variation in the reasons why interviewees had taken managerial positions in Case A. Like the majority of manager-academics at Case B, a few had taken the position reluctantly, with one department head upon joining the university being promised that they could do “professor’s work only, but that lasted about two months” (M-A 6, Case A). Others, some of who had occupied the department head position for long periods of time, hold the position to attempt to actively shape the development of their unit or department as well as the university more broadly. One interviewee mentioned that it was in his/her nature to want to influence the trajectory of his/her department.

I want to influence things. And if I see that things are not going in a good direction, I feel bad. So I have a kind of, maybe I have a feeling of responsibility, a kind of too big feeling, kind of, that's a personal characteristic, maybe. That's why I have been maybe involved so much in the development work over the years. I want to see things developing, I want to develop this university forward, have a strong kind of... And I think if I'm not involved, it's not happening. (M-A 3, Case A)

Another noted that they were “too conscious of the need to take care of some bigger issues” as they felt the need for “working for some greater good, to make things kind of work” within the changing environment in which the university now operates. But this is at great personal strain as they “seldom work the kind of normal office hours” (M-A 8, Case A).

Views on the role and responsibilities of a manager-academic in general

Despite these differences, department heads at both universities generally thought that the position enabled them to exert some degree of influence upon the academic agenda of their department. Manager-academics saw their capacity to directly affect the quality and quantity of research output, however, as limited. Rather their efforts were focused on influencing the working environment of the department, to provide the conditions conducive to academic success, and the necessary resources, such as administrative support and funding, to allow researchers to preserve their time for research.

I don't think the head of department can be directly related to research success but of course my duty is to provide a good environment and the necessary tools so that others can focus on their research. And by tools I mean, well, working environment, but also funding and encouragement and so I try to optimize these issues whenever possible. So in that sense maybe I'm responsible for and accountable for providing the necessary micro-cosmos so that others can perform well. (M-A 5, Case B)

One concern expressed by a number of manager-academics was the increasing bureaucratic burden imposed on departments and individual academics. When asked about what they intended to achieve in

their position as a department head, one interviewee from Case B replied that “I just try to act as a shield to the rest of the department [to] try to stop the bureaucracy somewhere, so it’s a defensive role” (M-A 8, Case B). Similarly, a department head at Case A commented that the ability of the department head to “buffer” staff from administrative tasks, in order to preserve time for academic work, is one of their primary responsibilities:

I think my major agenda is buffering our people. Nowadays it’s really possible for the department head to buffer a lot our researchers and teachers not needing to spend too much of their precious time to various kind of managerial activities. (M-A 4, Case A)

The majority of the department heads we interviewed, at both universities, saw development and evaluation discussion as a particularly important mechanism for shaping the academic activities of staff, especially for early-career researchers. Some regarded such formal mechanisms as being of little value, and adopted more informal approaches. One commented that providing guidance and feedback to staff “should be done every day at work [...] there’s no point in formalizing it” (M-A 5, Case B), while another saw their influence being exerted occasionally and informally by “just giving gentle nudges” when needed and by engaging in “genuine discussions about what [the staff] are doing” (M-A 7, Case B). This department head also commented that there is really little need for any formal processes for managing academics as long as they “have the right values” (M-A 7, Case B). Other department heads emphasized practices such as creating a culture around research seminar attendance and participation, visits by prominent international scholars, encouraging and facilitating collaboration within and cross departments, and providing feedback and guidance (especially to younger staff members). Such practices were common across the two cases.

Views on the local PMS and its implications

The top manager-academic of the faculty of Case A comments that with the system in place, department heads have a significant amount of autonomy in terms of how they improve academic performance, but reflects that the pressures from the field are such that there may not be any need for the points system, at least in terms of incentivizing academics to pursue top tier journal publications:

I think there is a lot of autonomy, and on the other hand trust, from my point of view in how [...] they then take care of the quality and quantity, it’s up to them. But we have to remember that at the same time there has been this institutional development going on over the years which we might call also kind of publish or perish mentality. So, especially the younger scholars know that they have to publish and in good journals if they want to have an academic career. So in that regard we don’t... sometimes I ask myself do we really need this point system or is this a kind of self-fulfilling prophecy or system that they will publish anyway in good journals. But that we don’t know. (M-A 2, Case A)

The department heads at Case A perceived a significant amount of autonomy in how they addressed academic performance in their department. However, the capacity to influence administrative structures in the school and wider university were seen to be diminishing.

My room for agency has diminished, yeah, that’s the fact [...] my agency in terms of what we see as important, developing the school and the research here, agency for that is broad and I could probably do

much more and be free to do that. But then in the everyday management, again, if we talk about management, there the room for agency has diminished somewhat. (M-A 3, Case A)

The diminished room for manager-academics' agency was far more evident, however, at Case B.

Substantial structural changes have seen a significant centralization of administrative and strategic decision-making, such as the introduction of the tenure track and narrow scope of performance evaluation. The top manager-academic of the faculty implies that given the current environment in which universities operate, there is some degree of inevitability about the changes:

This process, I don't pretend to be very important here. It is, I'm just a vehicle for this organization, the board, the top management. If I wouldn't be here, somebody else would be here and not doing identical work to what I'm doing, but there would be something in this, something like this happening here. (M-A 2, Case B)

Even though senior academics are nominally part of strategic discussions and decision-making, they have little capacity to actually influence these decisions unless they are a part of the initial process. Much of the decision-making at Case B is highly centralized, with the purpose of participation in dialogues to, in the words of the top manager-academic, "seduce" senior academics to buy into the changes.

Just because I'm a [the top manager], I can't make decisions that the professors don't like [...] I need to persuade them, I need to seduce them to, kind of, my side. Otherwise you can't accomplish these kinds of changes in a university. (M-A 1, Case B)

Department heads at Case B were generally of the opinion that their capacity to influence decisions of importance in the university are limited, and even intentionally bounded through centralization — participation is merely an illusion. One manager-academic gave an account of the decision-making process to relocate the department to a new campus that is in development. They noted that there was little transparency in the way the decision was determined, and there was the "impression that decisions were made long before professors are asked about their opinions. And we can express our opinions but they don't have any impact on the decisions because they have been previously made." (M-A 5, Case B). Another department head was even more scathing in assessing the administrative changes at the university:

Well they call it leaders' dialogue, but we call it leaders' monologue. They go around, they show a couple of slides, allow a couple of questions, and then go home and continue with the same set of slides. Feels like you are talking to a wall. (M-A 8, Case B)

Manager-academics' actions concerning research quality

At Case A there was a general view among manager-academics that they had to actively fight against the increasing pressures placed on academics to publish in the short-term. One department head saw a strong tension between the pressures to publish and the "compromising [of] research quality", especially in young academics:

This tendency to be rushing and instrumentalist, playing the publish or perish game. So that's the major tension, that's something I have to fight every day and it's close to every day, it's really all the time. (M-A 4, Case A)

Other manager-academics of Case A had similar sentiments.

People want to get their papers published quickly and I always say them that 'No, think about first what you write, think about where you submit, think about what does this paper mean to you after three, five years, not just now when you can put it in your CV. (M-A 3, Case A)

One of our researchers that has been in sort of an insecure position for some time, I've been all the time trying to mentor him to get rid of this idea that "I have to publish, I have to publish alone, I have to publish fast" but you should publish together, slow and good quality. (M-A 6, Case A)

At Case A, especially one head of department felt that determined and consistent long-term work for developing research seminar and discussion group culture is significant – and also pays off. This interviewee felt that leading by own example is of vital importance.

It is that I always talk about this: let's remember to think, let's invest a lot in the early process, let's not write before we are ready, let's write first the two pager or synopsis, let's talk to our colleagues about these ideas, these two pagers, let's even present them, let's discuss a lot, let's not start going and presenting these early papers in New Zealand or somewhere else first, let's go first to our seminar, let's develop a network to whom we can send papers for comments, let's pay back to those people who commented our papers, let's make it give and take in many ways in our group nationally, internationally, let's be active, work together. And I always try to show this model myself and talk about that and act like this in the seminars and discussion group meetings. So I normally always present my papers here at some point. (M-A 4, Case A).

But also others at Case A indicated quite strong vision and leadership type of agency:

IR: What do you want to achieve as an academic leader? Or is it more like fulfilling some expectations? Or both?

IE: [long silence] That's a tricky question. Of course because we come to the science itself and it's, kind of, let's say... [silence] I would like to, as an academic leader, I would like to kind of make an impact through-. I think I'm doing the greatest impact through my doctoral candidates. It's kind of trying to make them see things, to do innovative research, to do, see science as a broad... Let me think. As a broad and open-minded exercise. I hate narrow-mindedness, I hate only single-theory, only single-method, only single rules... That this kind of dogmatism is horror for me, so my kind of scientific ideal is that my researchers themselves realize this, that what would be an innovative perspective to deal with [the discipline of the interviewee], how could they [frustrated sound]... Only a little bit of... I also want to go a little bit against the mainstream, against the main wind, kind of, encouraging my people to do it a little bit differently and not to believe what the American elite says. (M-A 3, Case A).

At Case B, only one manager-academic noted any tension between the increasing pressure to publish in top ranked journals and research quality, in particular, the concern that academic performance at Case B is based almost entirely on a small set of journals (i.e., the GJRS1):

It has become more and more sort of this KPI driven type of thing so people need to publish, publish, publish, no matter whether the articles actually make sense or provide any contribution in the sense that they actually deal with interesting issues [...] I don't feel that we are actually in that sense getting to the right direction. I try to sort of fight against that but it seems a pretty hard battle. (M-A 4, Case B)

The idea to have GJRS1 as the benchmark for tenure-track academics, and further promotion potential, was largely imposed by the top management of the faculty. Department heads at Case B generally acknowledged that there are many discipline specific journals outside of the GJRS1, where high quality research can be, and is, published, but admitted the criterion does serve the purpose in signalling the strategic intentions of the university:

Taking only GJRS1 is far too limited. I understand that the ambition is to get people to understand that you have to aim at the highest level to be there and get the overall quality up. In that sense it communicates the expectations well but then again it's not a fair benchmark. (M-A 4, Case B)

The views of other department heads at Case B echoed this opinion. One stated that “there’s nothing in my power that I can do to squeeze GJRS publications” from their staff “except for saying ‘don’t submit to [discipline specific] journals’ [outside of the GJRS] which I’m not going to do” (M-A 7, Case B). Another department head lamented the increasing emphasis on using quantitative measures as the basis for performance evaluation as academic work is “also about the nontangible things and you cannot measure everything” (M-A 5, Case B). They understood, however, how it facilitates the bureaucratic management of the university:

I think it makes the lives of the decision makers easier by referring to KPIs and then everything is like very bureaucratic way [...] so you have then some kind of basis for your argumentation. But the truth is not as simple, it cannot be expressed in KPIs. It’s more complex than that. (M-A 5, Case B)

From previous experience at international universities, this department head considered that placing such substantial emphasis on essentially a single measure of performance can have significant consequences for the type of research that is pursued. They recall that at one particular institution, “senior people there proactively ask postdocs or tenure track people to eliminate projects which do not have this [highest level of] GJRS2 or GJRS1 potential” (M-A 5, Case B). Given the critical stance to performance measurement, this department head does not set any targets for their staff. Instead, they externalize this function to let “the market” (the academic community of their discipline) determine the incentives for researchers.

The market defines the targets [...] if you want to have a salary increase or a job, look at the market requirements. (M-A 5, Case B).

However, the requirements for promotion and increased compensation at Case B are essentially tied to publishing in the top tier publications as determined by the top management of the faculty, and everyone, more or less explicitly, understands this. Another head of a particularly successful department at Case B argued that there was an increasingly global convergence around the meaning of research quality and its assessment.

I think that nowadays actually there is more and more convergence towards, basically this kind of thinking and that’s what I like to think also. Meaning that there’s appreciation and understanding that good research can be of any kind as long as it’s ambitiously pursued and done seriously and one measure of that are the publications. (M-A 6, Case B)

When asked about how s/he guides young academics towards achieving quality research, the department head emphasized that they consider it a responsibility to ensure that they “first of all understand what the rules of the game in the world are, so I think that’s one starting point. Meaning that one has realistic expectations as to what can be accomplished and in what kind of time period” (M-A 6, Case B). S/he further elaborated that what does sometimes happen is that “people work towards their PhDs and then they are done and then suddenly they do encounter the challenges that [...] you should, you know, do this path breaking long term research and publish great papers and people may not be prepared for that” (M-A 6,

Case B). S/he makes the interesting comment that academia is “really a risk-taking business and it should be” (M-A 6, Case B) and this is something that is encouraged at the department.

In describing the research culture of the faculty, the top manager of Case B acknowledged that there was perhaps too much emphasis on achieving top tier publications:

I think the best way to put that would be to say that now, generally speaking, very high levels of ambition with perhaps a bit too much emphasis on publications in the top journals rather than quality per se. (M-A 2, Case B)

Another member of the top management team of the faculty at Case B also characterized the emphasis on “quality” research as being ingrained into the mindset of academics at the university.

I think that because our choice to try to get world class, the emphasis is all the time on quality. [...] Somehow it is present that people understand that we have to, kind of, produce quality, we have to aim higher, we have to improve. I think it is, kind of, built in to this university now. (M-A 3, Case B)

Yet the performance evaluation system associates quality with publication, and “especially in top journals”, to the exclusion of other measures: “if you think about our reward system, for example we don’t reward on citations. We reward on publications” (M-A 3, Case B). While the promotion prospects of more senior academics depend to some extent on thesis supervisions, external funding, and societal impact, publishing in the GJRS remains the dominant criterion. The top manager-academic of the university of Case B, in discussing recruitment decisions, even alludes to how this feeds into whether one is perceived to be a “good” academic:

Of course we look at that they have good papers, that they publish regularly and so on [...] things like that, but there are good papers where you can aim, these international GJRS1 and these kinds of things. So if you’ve managed to get many of those, then you’re considered a good professor. (M-A 1, Case B)

At Case A, there is significantly less emphasis on the use of performance measures to incentivize or evaluate academic performance. One department head, when asked about whether they use journal rankings to evaluate performance replied:

No, quality comes from the people and their orientation and understanding of the academic quality and what we try to achieve here. It doesn’t come from the managerial control systems. These are just checking that we are performing okay, fine in terms of numbers. That’s how I see it. And it’s good for that practice. But then the quality comes from other sources. (M-A 6, Case A)

This department head further commented that:

They all know that we should publish according to the NJRS classification but I do actually just the opposite, that I tell them that ‘yes it’s important but it’s not that important, it’s more important you bring about a credible research track as an individual and then you are able to take the others with you and work together and build a research community of your own’. I’m not sure if they are buying into this but that’s what I’m trying to tell them. (M-A 6, Case A)

Discussion

The meaning of quality in scholarly research

Our findings reveal, at least at a general level, a relatively consistent interpretation of the meaning of quality at the two case sites. Their interpretation of quality aligns with Karpik's (2010) notion of singularities. First, quality of research is multidimensional – for research to be of high quality there are multiple characteristics that it must display. These characteristics can be subsumed within two broad dimensions, which roughly correspond to Pirsig's "classic" and "romantic" ways of thinking about quality. In relation to the former, manager-academics noted that there were certain fundamental criteria that a study needs to comprise in order for it to be of quality – for instance, stating a clear research question, having a logical structure, rigorous theory development, using appropriate research methods and techniques, and adhering to relevant ethical standards. The other side of quality relates to the much more subjective and difficult to explicate aspect of research contribution – manager-academics variously referred to aspects such as novelty, originality, relevance, impact, and excitement (see Lamont, 2009). While there was general agreement amongst interviewees that high quality research must exhibit one or more of these characteristics, there was far less clarity around what they substantively constitute.

Second, research is often marked by uncertainty of quality in several ways. In particular, there can be a significant evaluation lag in forming a judgment of the quality of a piece of research, and the perception of quality of a piece of research can change over time, for instance, depending on the evaluating party (e.g., Lamont, . One manager-academic gave an example of a postdoc researcher who struggled to get funding for a research project, but this project ultimately led to him/her becoming a Nobel laureate.

Third, research is fundamentally incommensurable regarding quality. Certain 'conformance' criteria are relatively objective and definable in nature; a check-list of standards that must be fulfilled to reach a basic level of quality that is more or less expected to be achieved in any piece of scholarly research. Such standards are typically well established within a discipline, meaning that it is relatively easier to come to a consensus concerning the presence or absence of conformance criteria. But the nature of the transcendental or 'excellence' side of quality in research precludes a concrete definition. Because of this inherent incommensurability, any attempt to objectify it will inevitably capture "something less than Quality itself" (Pirsig, 2011, p. 236). Furthermore, the practical meaning of quality in research is potentially open to many interpretations – a piece of research may be seen as high quality by some but not by others. Nevertheless, consensus is often reached by disciplinary experts about research pieces that are, or are not, of high quality; but as this understanding is socially constructed within a particular research community, differences in quality may not be apparent, or easily explained, to an outsider.

Despite there being little variation between manager-academics on the conceptual meaning of research quality within and across the two case sites, there were differences in how it was employed in practice; these differences reflect different degrees of slippage between the ‘espoused-meaning’ of quality and its ‘meaning-in-practice’ (cf. Argyris & Schon, 1974). These differences can be traced to the ways judgment devices were mobilised, both those that originate within and outside the university, and the degree to which, and how, individual agency was exercised, by manager-academics at the two institutions.

Global and local judgment devices

Judgment devices are typically viewed as helpful mechanisms for individuals, and non-experts in particular, to make reasonable decisions involving research quality (Karpik, 2010). Paradoxically, judgment devices work by facilitating commensuration of the incommensurable – they tend to reduce or eliminate the need for expertise of an area of research or an understanding of the implicit values that underpin discipline-specific meanings of quality. Attempts to conduct such commensuration can make the forms and degree of heterogeneity between research outputs and their quality less visible.

Our particular focus in this study was on the local PMS, but invariably they have been influenced by significant changes occurring in the global and national context – the world of science is indeed global and ideologies related to university management seem to travel easily (cf. ter Bogt & Scapens, 2012; Watermeyer, 2014; Kallio et al., 2016). One of the most potent tendencies shaping the design and implementation of local PMSs is the growing importance of global university rankings. Consistent with Espeland and Stevens (1998), we find that despite many manager-academics openly criticising rankings, they are ultimately too influential to ignore. However, they have had somewhat different effects in the two cases. In Case A, the PMS is implemented in such a way that also publications in journals ranked below the top tier can constitute high quality research. The PMS also places a significant emphasis on citations, explicitly recognising that quality in terms of relevance and impact can only be fully appreciated over the longer term. The management ideology of Case B was explicitly grounded upon the notion of, and ambition to become, “world class”. Rather than being left as a nebulous concept, global ranking lists, and the criteria used in their construction, have come to constitute its meaning at Case B. The PMS at Case B thus reflects their vision of becoming world class – the PMS has been intentionally designed with a narrow focus on GJRS1 journals, as it is publications in this top echelon that matter most for increasing positional ranking.

An interesting difference between the two cases are perceptions relating to the different journal ranking systems that serve as a point of reference. While manager-academics at both case sites rely on judgment devices, they do that with different intensity and in different ways. Case A uses the NJRS, and although

there were some reservations, manager-academics generally felt that relative to other journal ranking schemes, the NJRS covered a wider sphere to disciplinary fields and took better into consideration journals that published also pieces following non-mainstream research paradigms. Hence, NJRS plays an important, yet not as such dominant, role in the PMS of Case A. In contrast, the NJRS classification was noted by some at Case B to be distorted, even “corrupted”. Many manager-academics at Case B believed quite firmly in the GJRS as a kind of objective and commonly agreed upon measure of research quality. As such, evaluation is essentially outsourced to a third-party which is not an academic institution nor fully transparent regarding how the GRJS is constructed, while the NJRS, deliberated upon by a panel of academics, that also builds upon bibliometric analysis, is straightforwardly considered to be a less valid and reliable measure of quality. The role played by GJRS in the performance evaluation of scholars at case B was very big and dominant.

Manager-academics in both case sites acknowledged that objective indicators, i.e., various kinds of judgment devices, were often drawn upon to make assessments about the quality of research produced by an academic – e.g., tenure track evaluations, hiring selections, and staff promotions. In Case A, the PMS was primarily seen as a mechanism for self-reflection and a basis for development discussions, and more explicit acknowledgement that conducting quality research may take a long time and not necessarily be published in the highest ranked journals. There were relatively loose ties between the PMS and tenure-track and promotion decisions, and publications in different NJRS tiers were associated with compensation that served more as a symbolic token than an explicit motivating factor. At Case B the PMS was linked to strong financial incentives, the ability to secure tenure, and the likelihood of promotion — this made it clear that publishing in GJRS journals is not only expected, but a requirement for long-term employment and career development (cf. van Dalen & Henkens, 2012). These reinforcing mechanisms created less space for individual interpretation in making judgments about research quality and what it is to be a ‘good scholar’ (cf. Hopwood, 2007; Czarniawska, 2011; Alvesson et al., 2017).

Differences between the two institutions should not be overstated – for instance, publications in high ranked journals are still an important marker of research quality at Case A, too. But despite many acknowledging that it is an imperfect measure of quality, there was a greater tendency for manager-academics at Case B to defer to PMS – together with its ‘outsourcing agents’ like GRJS – as the primary judgment device. This observation is consistent with prior research on surrogation (the process of replacing an underlying construct with an objective measure), which is more likely to occur precisely when the underlying construct is “abstract, ill-defined and complex” (i.e., quality of research) and individuals are incentivized on a single performance measure (Choi, Hecht & Tayler, 2012, p. 1141). Ultimately, the

unidimensional nature of the PMS and its more rigid application in determinations of research quality at Case B increased the divergence between the espoused meaning of quality and its translation in practice.

Manager-academics' agency

At both cases, manager-academics considered quality as fundamental to academic scholarship, yet few actively attempted to shape its meaning in practice. Manager-academics generally perceived their capacity to influence research quality as limited, such that they focused their efforts towards providing conditions that make it easier for academics to focus on research activities – for example, fostering a conducive work environment, buffering from administrative tasks, and making sufficient resources available. Most manager-academics also perceived their agentic capacity diminished – pressures from global competition and changes to national funding arrangements induced a feeling of inevitability regarding changes to university management and the evaluation of academics and their research.

Nevertheless, albeit not overly pronounced, a pattern emerged between the two cases regarding the perceived potential of, and need to exercise, agency. Manager-academics at Case B tended to hold one of two general positions. The more common was the view that if academics have the correct values then they can be left to their own devices — they are so self-motivated that there is essentially no need for any managerial interventions or academic leadership (e.g., M-A 4, Case B). The meaning of quality is taken-for-granted – it is “built into” the university (M-A 3, Case B). The other position was that determinations of research quality should be done by “the market.” The relevant academic community provides judgments through seminars, conferences, and ultimately through the journal review process.

Both positions appearing at Case B represent instances of manager-academics lending their agency away, trusting on either academics' self-governance or the academic field to define and make judgments on the quality of research, and both are problematic when it comes to safeguarding the incommensurability of research quality. The danger with leaving academics to their own devices is that they will be influenced, subtly and over time, to conform to increasing pressures, both from within and beyond the university, that shape the meaning of quality and the nature of the research that they do. Even though manager-academics at Case B perceived the local PMS as overly narrow, most either did not see the need, or were reluctant, to stage any “acts of resistance” (cf. Anderson, 2008; Kalfa et al., 2018). They were willing to dedicate only so much time to their managerial duties – after all, they are subject to the same publication pressures as those beneath them – leaving the effects of the local PMS on the practical meaning of research quality unmitigated. This is consistent with the argument of Espeland and Stevens (1998) that as commensuration is cemented within the practices and processes of organizations “it becomes more taken for granted and more constitutive of what it measures” (p. 329).

For one head of department (M-A 5, Case B), the entire PMS exercise is only “theatre”, in which all parties just pretend to be acting seriously. Only what “people in your field” think, is relevant. However, the global level the conception of high quality among the academics in the field is not that dissimilar to the local PMS at Case B – as another manager-academic suggested, there is increasing convergence around what constitutes good research in the academe and publications are representative of the level of quality (M-A 6, Case B). This is also shared by a senior-manager academic who has had significant influence on the design of the PMS (M-A 2, Case B). In effect, this has been an agentic decision to reproduce, rather than challenge, the global performance measurement regime, in an effort to become a “world class” university. Hence, as such at Case B, there is not that much difference in the outcomes as to whether manager-academics place their trust in autonomous self-governance or judgments by the field.

Manager-academics at Case A demonstrated more of an agentic tendency to feel a need to fight against the short-term pressures to perform. Especially M-A 4 of Case A presented strong opinions concerning the need to counteract against “playing of the game” towards publications, noting their active engagement with academics with the department to guide them away from such inclinations. At least for this manager-academic, the growing influence of commensuration through journal rankings made the defence of quality scholarship more meaningful and their defence more necessary (cf. Espeland & Stevens, 1998). A few other manager-academics at Case B (M-A 2, M-A 3, M-A 4, M-A 6) also held the belief that quality – and without defining it strictly and passively on a heteronomous basis – can be influenced through consistent and determined long-term series of efforts, by developing a collaborative culture, frequent interactions with academics, and leading by example. Similar views were expressed by M-A 4 from Case B, too, but s/he indicated less need or willingness to provide leadership in this regard.

At Case A, manager-academics also saw the main threat to research quality arising from the changing global and national context of academia, rather than the local PMS. The multi-dimensional PMS of Case A could in fact be mobilised to support an alignment between the espoused meaning of quality and its meaning in practice, as it encapsulated multiple performance measures to reflect quality and looser ties to formal performance and promotion assessments. This provided greater space for manager-academics to exert some degree of agency in defending the incommensurability of research quality. For example, the challenges related to publishing inter-disciplinary research in top-tier journals (e.g., in the largely discipline-focused GJRS journals), were considered seriously. But while there was certainly more concern about the influence of pressures from beyond the local context on research quality at Case A, we still observed only a few manager-academics making intensive and concerted efforts to alleviate them.

Overall, while certainly the importance of quality is broadly stressed, securing quality did not emerge prominently as a clear target for manager-academics' agency. In both cases, quality is typically left to be 'automatically' taken care of by the self-governance of the academics, by the expectations from the field, or by the PMSs. The general picture is hence that of a passive (rather than active) and heteronomous (rather than autonomous) position adopted by manager-academics (cf. Karpik, 2010). Many manager-academics, albeit often reluctantly, took 'short-cuts' in making judgmental evaluations, by relying on journal rankings (like the NJRS or GJRSs) and other measures captured by the local PMS. However, the employed reference dimensions for heteronomy is notably broader at Case A than at Case B, due to the more multi-dimensional PMS and the looser connection to decisions that involve determinations of research quality (e.g., tenure, promotion). This makes heteronomous judgment more inclusive at Case A than at Case B. At Case A, there is more perceived room for manoeuvre for manager-academics and somewhat more spelled out need and use of autonomous scholarly judgment. Some insisting on the need for active judgment, rather than just leaning on journal rankings surfaced also at Case B, but this seemed to be more in principle than actual practice. The overall findings in our analysis supports the view that it is typical of academics to not be particularly conscious about their agentic opportunities or the need to exercise them (cf. Alvesson et al. 2017).

Practical implications

Our findings generally indicate how the romantic notion of quality is increasingly squeezed into a narrow space in the realm of academic practice. Our findings generally support the view that passive and heteronomous types involvement of academics regarding assessing quality – thereby leaning increasingly on the use of various kinds of judgment devices – is gaining more ground, while active and autonomous involvement seems to be a losing battle – yet not without qualifications. While Karpik (2010) argues that commensuration enabled by judgment devices is not a threat to incommensurability as long it is “pluralist and reversible” (p. 12) and “safeguards the diversity of personal interpretations” (p. 18), such securing of incommensurability is easily under great threat in the current circumstances of the publish or perish culture and managerialism. This shows very explicitly at Case B, where a straightforward commensuration of research quality has, due to its narrow PMS, tightly coupled to the tenure track system of the unit, become *de facto* the default. This also shows in the reluctance of many manager-academics to even hold their positions at Case B: They feel they are only a part of the administrative apparatus and many of them feel such work is just wasting their precious time. The feeling of agency as manager-academics is generally weak at Case B. At that case site, we might suggest that a vicious circle has emerged in which the dysfunctional effects – at least regarding the romantic notion of quality – of instrumentalism relating to the publish or

perish culture have become emphasized. Managerialist take to running the faculty, which leans notably to a passive and heteronomous use of judgments devices, has taken over.

But even at Case A, where the PMS is notably broader and less coupled to career decisions, a commensurating tendency regarding manager-academics' practices related to research quality is looming and may be in a process of become stronger. Also at Case A, everybody is busy and the short-cuts offered by various kinds of 'handy' judgment devices in various kinds of performance measurement and management decision situations feel often persuasive. As their use is also fashionable and is hence encountered in all kinds of instances both nationally and internationally, their use has an 'easy-ride' (Granlund et al., 1998). It is easy for all to forget the other alternative: active and autonomous judgment, with only small or moderate reliance of the 'modern' judgment devices like journal rankings or impact factors. After all, also academics tend to be flock-oriented humans. However, the PMS of Case A, purposefully designed to be broad and inclusive, still makes a notable difference: There is a significantly bigger room for manager-academics' agency at Case A than at Case B. Several manager-academics at Case A also exert their agency, even though the extent to its exertion differs. Practising agency with a view of allowing room for the romantic notion of quality, too, is anyhow much more possible at Case A than at Case B. This suggests that it matters how the local PMS is designed and used in an academic unit, since the agency of manager-academics (and academics overall) necessarily requires sufficient room for manoeuvre.

Conclusions

In this study we examined the practical meaning and employment of quality in scholarly research from the perspective of manager-academics, using comparative case study materials from two university faculties. We were concerned with how both local and global pressures are influencing how manager-academics perceive the notion of quality in research and how this notion is employed and becomes embedded in local practices. In particular, we paid close attention to whether, and if so how, academic-managers exerted their agency to promote or buffer academics from these pressures, and the implications that this has for what quality in research is understood to be. The broad underlying motivation for our analysis was our worry of the potential disappearance of signs of, and room for, the romantic notion of quality in the academic practice due to the likely increasing use of various kind of judgment devices offering easy short-cuts for working around research quality. In this way academic practices around research quality would be at risk of becoming just technical, almost secretarial tasks (cf. Hopper, 2016).

The faculties we examined face the same global and national pressures, and both are pursuing agendas to become 'world class' institutions. The two cases, however, have adopted somewhat different approaches

to performance measurement. Invariably, publications matter, in terms of quantity and the outlet, but more so in Case B where the PMS is almost exclusively focused on publications in a narrow 'qualified' set of journals. In Case A the PMS is much more inclusive as to what journals academics are incentivized to publish in, and it casts a wider net around what is included in judgmental evaluations of academic performance. Despite these differences the general understanding of the notion of quality in research is relatively consistent across manager-academics at the two cases. This understanding broadly correlates to the two ways of thinking about quality presented by Pirsig (1974) – the classic and the romantic. Both are necessary components for a piece of research to be deemed as high quality. But because the romantic element of quality is inherently subjective, intangible, and difficult to determine, it is more likely to be marginalised in attempts to capture quality through measurable proxies. This leads to a tendency of a narrowing meaning of quality in actual practice, which we also find in our two case faculties, yet to differing degrees.

We expected that manager-academic agency would be an important mechanism through which global and local pressures would be mediated, and a sufficient degree of the romantic and idealistic understanding of quality and scholarship could be safeguarded. While we did observe instances of manager-academics demonstrating strong agency in this regard – and notably more so at Case A than at Case B – it was more the exception than the norm. Manager-academics were more likely to simply reproduce global pressures locally, that is, accepting that this is simply the way the academic world now operates, or not exercising their agency actively, trusting the self-governing capacities of individual academics, and relying on the 'market forces' of the academic community to determine what constitutes quality in research. That said, both the ability and willingness to exert agency seemed more limited in Case B than Case A, which seems to be at least partly attributable to the more restrictive PMS in place in the former mentioned faculty. Consequently, at Case B there is little incentive to take the role of a manager-academic, particularly as a department head, as they are subject to the same system of evaluation as academics without managerial responsibilities. Hence it is little surprise that, even if somewhat reluctantly, manager-academics at Case B, but to a lesser extent those at Case A, rely upon various kinds of technical performance measures (i.e., judgment devices) to make evaluations about research quality.

There are limitations to this study. Although we took into consideration structural and cultural factors in our two case sites, our focus was largely on the role of PMS, and as such, we provide only a partial understanding of how variations in performance management practices influence the meaning and employment of quality in research. We also examine two faculties in a single national context. Although institutional and regulatory changes mirror much of those that have occurred in other Western nations, there may be cultural and other factors that shape the way that manager-academics in our study respond to intensifying pressures, both managerial and those from 'the field'. Despite these limitations, we hope

that this study sheds light on the meaning of quality in scholarly research and how this notion is being shaped by global and local pressures.

We conclude, firstly, that manager-academic agency is important, but infrequently mobilised, for safeguarding the romantic meaning of quality in research. Without this aspect of thinking and practising quality in the academe, there is a high risk of its excessive commensuration, likely in the long run leading to seriously lower degrees of freedom and hence heterogeneity in research topics and approaches (Lukka & Mouritsen, 2002). Quality assessment should not be viewed as only a technical operation, taking place straightforwardly through the fashionable and 'handy' judgment devices. Our second conclusion is that manager-academics' agency, for the same purpose, requires a sufficiently broad and inclusive local PMS to offer the necessary room for manoeuvre. Hence, we argue that ever more so as the global publish-or-perish culture is getting increasingly intensive and pressing, the way in which the local PMS of academic units are designed and used really matters. Hence, while a broad and open-minded understanding and employment of the notion of quality in research – fundamentally at least partly an 'unknown' thing – is at constant danger these days, there is still hope. And what is important to realise, it is not only an exogenous systemic threat, but rather notably an endogenous one: It is up to the academics themselves.

References

- Adler, N. J., & Harzing, A. W. (2009). When knowledge wins: Transcending the sense and nonsense of academic rankings. *Academy of Management Learning & Education*, 8(1), 72–95.
- Ahrens, T., & Chapman, C. S. (2006). Doing qualitative field research in management accounting: Positioning data to contribute to theory. *Accounting, Organizations and Society*, 31(8), 819–841.
- Alvesson, M. (2003). Methodology for close up studies – struggling with closeness and closure. *Higher Education*, 46(2), 167–193.
- Alvesson, M., Gabriel, Y., & Paulsen, R. (2017). *Return to Meaning: A Social Science with Something to Say*. Oxford University Press.
- Archer, M. S. (2007). *Making our way through the world: human reflexivity and social mobility*. Cambridge: Cambridge University Press.
- Bachrach, D. G., Bendoly, E., Beu Ammeter, D., Blackburn, R., Brown, K. G., Burke, G., Callahan, T., Chen, K. Y., Day, V. H., Ellstrand, A.E., & Erekson, O.H. (2017). On academic rankings, unacceptable methods, and the social obligations of business schools. *Decision Sciences*, 48(3), 561–585.
- Ball, C. (1985). What the hell is quality. *Fitness for Purpose*, 96–102.
- Bialecki, M., O’Leary, S. and Smith, D., 2017. Judgement devices and the evaluation of singularities: The use of performance ratings and narrative information to guide film viewer choice. *Management Accounting Research*, 56-65.
- Broadbent, J., Jacobs, K., & Laughlin, R. (1999). Comparing schools in the UK and New Zealand: Individualizing and socializing accountabilities and some implications for management control. *Management Accounting Research*, 10(4), 339–361.
- Brunsson, N. (1985). *The Irrational Organization. Irrationality as a Basis for Organizational Action and Change*. New York, NY: Wiley.
- Choi, J., Hecht, G. W., & Tayler, W. B. (2012). Lost in translation: The effects of incentive compensation on strategy surrogation. *The Accounting Review*, 87(4), 1135–1163.
- Czarniawska, B. (2011). Successful Research: In Whose Eyes? *European Accounting Review*, 20(1), 53–55.
- Deem, R. (1998). 'New managerialism' and higher education: The management of performances and cultures in universities in the United Kingdom. *International Studies in Sociology of Education*, 8(1), 47–70.
- Deem, R. (2004). The knowledge worker, the manager-academic and the contemporary UK university: New and old forms of public management?. *Financial Accountability & Management*, 20(2), 107–128.
- Dubois, A., & Gadde, L-E. (2002) Systematic combining: An abductive approach to case research. *Journal of Business Research*, 55(7), 553–560.

- Durand, R., & Vaara, E. (2009). Causation, counterfactuals and competitive advantage. *Strategic Management Journal*, 30(12), 1245–1264.
- Edwards, M. A., & Roy, S. (2017). Academic research in the 21st century: Maintaining scientific integrity in a climate of perverse incentives and hypercompetition. *Environmental Engineering Science*, 34(1), 51–61.
- Endenich, C., & Trapp, R. (2018). Signaling effects of scholarly profiles—The editorial teams of North American accounting association journals. *Critical Perspectives on Accounting*, 51, 4–23.
<http://dx.doi.org/10.1016/j.cpa.2017.02.004>
- Enders, J. (2004). Higher Education, internationalisation, and the nation-state: Recent developments and challenges for governance theory. *Higher Education*, 47(3), 361–382
- Garvin, D. A. (1984). What does “product quality” really mean? *Sloan Management Review*, Fall, 25–43.
- Gendron, Y. (2008). Constituting the academic performer: The spectre of superficiality and stagnation in academia. *European Accounting Review*, 17 (1), 97–127.
- Gendron, Y. (2013). Rethinking the concept of research contribution. *Comptabilité-Contrôle-Audit*, 19(2), 133–155.
- Gendron, Y. (2015). Accounting academia and the threat of the paying-off mentality. *Critical Perspectives on Accounting*, 26, 168–176.
- Giacalone, R. A. (2009). Academic rankings in research institutions: A case of skewed mind-sets and professional amnesia. *Academy of Management Learning & Education*, 8(1), 122–126.
- Giddens, A. (1984). The constitution of society. Outline of the theory of structuration. Cambridge: Polity.
- Graf, L., Wendler, W.S., Stumpf-Wollersheim, J. & Welp, I.M. (2017). Wanting more, getting less: Gaming performance measurement as a form of deviant workplace behavior. *Journal of Business Ethics*, 1-21.
- Granlund, M., Lukka, K., and Mouritsen, J. (1998) Institutionalised justification of corporate action: internationalisation and the EU in corporate reports. *Scandinavian Journal of Management*, 14(4): 433-458.
- Greenwood, R., Díaz, A. M., Li, S. X., & Lorente, J. C. (2010). The multiplicity of institutional logics and the heterogeneity of organizational responses. *Organization Science*, 21(2), 521-539.
- Gross, P. R. & Levitt, N. (1994). *Higher Superstition: The Academic Left and Its Quarrels With Science*. Baltimore: Johns Hopkins University Press.
- Guthrie, J., & Neumann, R. (2007). Economic and non-financial performance indicators in universities: The establishment of a performance-driven system for Australian higher education. *Public Management Review*, 9(2), pp.231–252.
- Hacking, I. (1999). *The social construction of what?* Harvard University Press.
- Harvey, L., & Green, D. (1993). Defining quality. *Assessment & Evaluation in Higher Education*, 18 (1), 9–34.

- Hood, C. (1995). The “New Public Management” in the 1980s: Variations on a theme. *Accounting, Organizations and Society*, 20(2-3), 93–109.
- Hopwood, A. G. (2007). Whither accounting research? *The Accounting Review*, 82(5), 1365–1374.
- Hopper, T. (2016) [Editorial] *Accounting scholarship and management by numbers*. Revista Contabilidade & Finanças (Accounting and Finance Review), 27 (71). pp. 144-148. ISSN 1519-7077.
- Humphrey, C., & Genrdon, Y. (2015). What is going on? The sustainability of accounting academia. *Critical Perspectives on Accounting*, 26, 47–66.
- Jordan, S., & Messner, M. (2012). Enabling control and the problem of incomplete performance indicators. *Accounting, Organizations and Society*, 37, 544–564.
- Jönsson, S., & Lukka, K. (2006). There and back again: Doing interventionist research in management accounting. In C. S. Chapman, A. G. Hopwood, & M. D. Shields (Eds.), *Handbook of management accounting research* (Vol. 1, pp. 373–397). Amsterdam: Elsevier.
- Kallio, K. M., & Kallio, T. J. (2014). Management-by-results and performance measurement in universities—implications for work motivation. *Studies in Higher Education*, 39(4), 574–589.
- Kallio, K. M., Kallio, T. J., Tienari, J., & Hyvönen, T. (2016). Ethos at stake: Performance management and academic work in universities. *Human Relations*, 69(3), 685–709.
- Kallio, K. M., Kallio, T. J., & Grossi, G. (2017). Performance measurement in universities: Ambiguities in the use of quality versus quantity in performance indicators. *Public Money & Management*, 37(4), 293–300.
- Karpik, L. (2010). *Valuing the unique. The economics of singularities*. New Jersey: Princeton University Press.
- Lamont, M. (2009). *How professors think*. Harvard University Press.
- Lamont, M. & Guetzkow, J. (2016). How Quality Is Recognized by Peer Review Panels: The Case of the Humanities. In *Research Assessment in the Humanities* (pp. 31-41). Springer, Cham.
- Lapsley, I. (1999). Accounting and the new public management: instruments of substantive efficiency or a rationalising modernity?. *Financial Accountability & Management*, 15(3–4), 201–207.
- Lewis, J. M. (2014). Research productivity and research system attitudes. *Public Money & Management*, 34, 6, 417–424.
- Lukka, K. (2014). Exploring the possibilities for causal explanation in interpretive research. *Accounting, Organizations and Society*, 39(7), 559–566.
- Lukka, K., & Modell, S. (2010). Validation in interpretive management accounting research. *Accounting, Organizations and Society*, 35(4), 462–477.
- Lukka, K. & Mouritsen, J. (2002). Homogeneity or heterogeneity of research in management accounting?. *European Accounting Review*, 11(4), 805-811.
- Macdonald, S., & Kam, J. (2007). Ring a ring o’roses: Quality journals and gamesmanship in management studies. *Journal of Management Studies*, 44(4), 640–655.

- Macintosh, N. B. (2004) A ghostly CAR ride. *Critical Perspectives on Accounting*, 15(4), pp.675–695.
- Martin-Sardesai, A., Irvine, H., Tooley, S., & Guthrie, J. (2017). Government research evaluations and academic freedom: A UK and Australian comparison. *Higher Education Research & Development*, 36(2), 372–385.
- Melan, M. (2013) *Will and mechanisms*. Reports from the Department of Philosophy. University of Turku.
- Modell, S. (2003), Goals versus institutions: The development of performance measurement in the Swedish university sector. *Management Accounting Research*, 14, 4, 333–359.
- Modell, S. (2005). Students as consumers? An institutional field-level analysis of the construction of performance measurement practices. *Accounting, Auditing & Accountability Journal*, 18(4), 537–563.
- Modell, S. (2017). Critical realist accounting research: In search of its emancipatory potential. *Critical Perspectives on Accounting*, 42, 20–35.
- Neumann, R., & Guthrie, J. (2002). The corporatization of research in Australian higher education. *Critical Perspectives on Accounting*, 13(5–6), 721–741.
- Parker, L., & Gurthrie, J. (2013) Accounting scholars and journal rating and benchmarking. *Accounting, Auditing and Accountability Journal*, 25 (1), 4–15.
- Parker, M., & Jary, D. (1995). The McUniversity: Organization, management and academic subjectivity. *Organization*, 2(2), 319–338.
- Pirsig, R. M. (1974). *Zen and the Art of Motorcycle Maintenance. An Inquiry into Values*. Bodley Head.
- Prichard, C., & Willmott, H. (1997). Just How Managed is the McUniversity?. *Organization Studies*, 18(2), 287–316.
- Reeves, C. A., & Bednar, D. A. (1994). Defining quality: Alternatives and implications. *Academy of Management Review*, 19 (3), 419–445.
- Sauder, M. and Espeland, W., (2006) Strength in Numbers? The Advantages of Multiple Rankings. *Indiana Law Journal*, 81, 205-227.
- Shields, P. M. (1999). Zen and the art of higher education maintenance: Bridging classic and romantic notions of quality. *Journal of Higher Education Policy and Management*, 21 (2), 165–172.
- Sousa, C. A., de Nijs, W. F., & Hendriks, P. H. (2010). Secrets of the beehive: Performance management in university research organizations. *Human Relations*, 63(9), 1439–1460.
- Stones, R. (2005). *Structuration Theory*. London: Palgrave.
- Sutton, N. C., & Brown, D. A. (2016). The illusion of no control: management control systems facilitating autonomous motivation in university research. *Accounting & Finance*, 56(2), 577–604.
- ter Bogt, H.J., & Scapens, R.W. (2012). Performance management in universities: Effects of the transition to more quantitative measurement systems. *European Accounting Review*, 21(3), 451–497.

- Thornton, P. H., & Ocasio, W. (1999). Institutional logics and the historical contingency of power in organizations: Executive succession in the higher education publishing industry, 1958–1990. *American Journal of Sociology*, 105(3), 801–843.
- Thornton, P. H., Ocasio, W., & Lounsbury, M. (2012). *The Institutional Logics Perspective: A New Approach to Culture, Structure, and Process*. Oxford: Oxford University Press.
- Townley, B. (1997). The institutional logic of performance appraisal. *Organization Studies*, 18(2), 261–285.
- Vakkuri, J., & Meklin, P. (2003). The impact of culture on the use of performance measurement information in the university setting. *Management Decision*, 41 (8), 751–759.
- van Dooren, W., & Thijs, N. (2010). Paradoxes of improving performance management (systems) in public administration. *EIPAScope*, 2, 13–18.
- Van Dalen, H. P., & Henkens, K. (2012). Intended and unintended consequences of a publish-or-perish culture: A worldwide survey. *Journal of the Association for Information Science and Technology*, 63(7), 1282–1293.
- Watermeyer, R. (2014). Issues in the articulation of ‘impact’: the responses of UK academics to ‘impact’ as a new measure of research assessment. *Studies in Higher Education*, 39(2), 359–377.
- Willmott, H. (1995). Managing the academics: Commodification and control in the development of university education in the UK. *Human Relations*, 48(9), 993–1027.
- Willmott, H. (2011). Journal list fetishism and the perversion of scholarship: Reactivity and the ABS list. *Organization*, 18 (4), 429–442.
- Wittek, L., & Kvernbekk, T. (2011). On the problems of asking for a definition of quality in education. *Scandinavian Journal of Educational Research*, 55 (6), 671–684.
- Wiittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Blackwell.
- Yudkevich, M., Altbach, P.G., & Rumbley, L. E. (2016). Citius, altius, fortius: Global university rankings as the “Olympic Games” of higher education? *International Higher Education*, 84, 4–6.