

## Advancing to the Next Level: Caring for Evaluative Metrics Monsters in Academia and Healthcare

Iris Wallenburg, Wolfgang Kaltenbrunner, Björn Hammarfelt, Sarah Rijcke, Roland Bal

► **To cite this version:**

Iris Wallenburg, Wolfgang Kaltenbrunner, Björn Hammarfelt, Sarah Rijcke, Roland Bal. Advancing to the Next Level: Caring for Evaluative Metrics Monsters in Academia and Healthcare. Working Conference on Information Systems and Organizations (IS&O), Dec 2018, San Francisco, CA, United States. pp.80-95, 10.1007/978-3-030-04091-8\_7. hal-02083588

**HAL Id: hal-02083588**

**<https://hal.inria.fr/hal-02083588>**

Submitted on 29 Mar 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Advancing to the Next Level: Caring for Evaluative Metrics Monsters in Academia and Healthcare

Iris Wallenburg<sup>1</sup>[0000-0002-3132-4628], Wolfgang Kaltenbrunner<sup>2</sup>[0000-0003-4311-8212], Björn Hammarfelt<sup>3</sup>[0000-0002-1504-8395], Sarah de Rijcke<sup>4</sup>[0000-0003-4652-0362] and Roland Bal<sup>5</sup>[0000-0001-7202-5053]

<sup>1</sup> Erasmus School of Health Policy and Management, Rotterdam, The Netherlands, wallenburg@eshpm.eur.nl

<sup>2</sup> Centre for Science and Technology Studies, Leiden, The Netherlands, w.kaltenbrunner@cwts.leidenuniv.nl

<sup>3</sup> Swedish School of Library and Information Science, Borås, Sweden, bjorn.hammarfelt@hb.se

<sup>4</sup> Centre for Science and Technology Studies, Leiden, The Netherlands, s.de.rijcke@cwts.leidenuniv.nl

<sup>5</sup> Erasmus School of Health Policy and Management, Rotterdam, The Netherlands, r.bal@eshpm.eur.nl

**Abstract.** In this paper we use the notions of play and (finite and infinite) games to analyze performance management practices in professional work. Whilst evaluative metrics are often described as ‘monsters’ impacting on professional work, we illustrate how metrics can also become part of practices of caring for such work. Analyzing the use of evaluative metrics in law faculties and in hospitals, we show how finite games – games played to win – and infinite games – games played for the purpose of continuing to play – are intertwined and how this intertwinement affects academic and healthcare work.

**Keywords:** gamification, performance management, universities, health care

## 1 Introduction

Academic and healthcare life have increasingly metricized in the past two decades. Citation indexes, teaching quality assessment, research assessment, and league tables – and their underlying algorithms – have set a new scene for academic work and academic career development [1, 2]. Likewise, healthcare professionals face a myriad of different metrics, ranging from performance indicators for measuring and comparing quality of care, clinical registries to monitor patient outcomes and medical decision-making, and training quality evaluation, to instruments to measure patient values and establish cost effectiveness ratios for clinical treatment [3, 4].

A growing body of (STS) literature discusses how evaluative metrics transform professional work as they value certain worths over others, and shape what gets obscured, what gets highlighted and what is conceived of as good performance – thoroughly changing how organizations and individual practitioners shape their services and how they compete [5-9]. In academia, metrics like the h-index characterize and value the

scientific output of the individual researcher, stressing the importance of individual excellence and drawing attention away from collective work [10]. Also, the focus on publishing in (highly rewarding) high-impact journals displaces the importance of books and edited volumes, particularly those in a non-English language [11]. Kelly & Burrows [1] have stated that “(...) the enactment of value and relative worth in academic work by formal processes of academic judgment, measurement and algorithmic resource allocation has become fundamental to survival” – indicating how performance metrics actively reshape academic fields and their (e)valuation infrastructures [e.g. 5, 12]. Such conclusions often reflect the (private) worries of scholars about the impact of performance metrics on both their beloved academic fields and their individual careers, especially concerning the epistemic changes that metrics make happen, valuing specific kinds of knowledge and specific types of research over others. Healthcare practitioners express similar concerns. They argue that the increasing emphasis on performance measurement and control [e.g. 3, 4, 13], causes paper work and ‘red tape’ [14] stealing away time from their real jobs – that is, curing and caring for patients. Hence, metrics change healthcare work through emphasizing measurable aspects of care over others.

Yet, professionals, scientists and physicians alike, are not just victims of performance management but have played an active part in their construction. Universities themselves have crafted the first university ranking and evaluative instruments to pin down academic performance, and they have continued to do so [e.g. 15]. Similarly, medical associations have played a crucial role in designing performance indicators and evaluative instruments for setting clinical norms for good care and comparing clinical outcomes. They closely collaborate with other stakeholders (e.g. healthcare authorities, third party payers, patient associations) in defining performance instruments and outcome measurements [16, 17]. Hence, while practitioners groan under evaluative pressures, they have similarly embraced evaluative metrics that (re)shape their work. In medicine, for instance, physicians actively adopt value-based healthcare instruments to design patient-centered treatment trajectories, simultaneously establishing new and ambitious research lines and hence prompting new career opportunities [18]. This seems to point at another direction than that usually taken by the literature on audit and accountability, and allows for a conceptualization of metrics in more open ways, considering how metrics are used and play out in unfolding professional worlds. Empirical findings suggest that evaluative metrics allow for more diverse and ‘playful’ ways of evaluating and assessing professionals [8, 19]. Online tools, for instance, may also liberate scholars as they can profile themselves and their research to a wider public (or, more strategically, to a specific public in case of for instance a grant application) through building an online research identity [20].

This double-sided view on evaluative metrics is the central concern of this paper: evaluative metrics are seen as monsters threatening the ‘real work’. They would evoke practices of ‘gaming’ in which professionals seek to ‘cheat’ with numbers in order to cosmetically comply with them while actually continuing their own practices, as Bevan and Hood have strikingly pointed out as “hitting the target but missing the point” [21]. Yet, as we claim in this paper, evaluative metrics are also cared for, actively used and played with, prompting a more reflexive approach on performance – also among practitioners that can be quite cynical about them. So, what if we consider performance

metrics as a matter of play, encompassing both matters of concern [22] and matters of care [23]?

In this paper, we apply an experimental and explorative view on evaluative metrics, asking: How do academic and healthcare professionals playfully employ evaluative metrics in constituting academic/care work and professional routines, as well as to the purpose of building a professional position? With what emerging consequences for professional identity construction?

## 2 Playing the metrics game

In *Homo Ludens*, Johan Huizinga portrays the human being as intrinsically playful, and positions ‘play’ as a serious and necessary form of social interaction and cultural development [24]. Play, Huizinga argues, is something common and contributes to human and cultural development through modulation of experience [25]. Playing is a medium where lived experience is organized as a structured situation. It is something outside ‘normal’ life; play is spatial-temporal as playing is bounded to a well-defined situation, comprising clear rules. Play, Huizinga contends, expands from early childhood to old age. It can be both fun and serious and may also encompass suffering. Huizinga situates play in children living their own phantasies at the playground, but also in religious (and sometimes cruel) rituals, in poetry, art, science, the law, and sports.

Although Huizinga demarcates play from ordinary life, at the same time he acknowledges that boundaries between playing and non-playing can be permeable. For example, people may play with one another over dinner slipping in and out of serious interaction throughout an entire evening [25]. Stressing the complexity of separating the two in contemporary times he states (written in 1938): “it [the civilization] plays false, so that it becomes increasingly difficult to tell where play ends, and non-play begins” (25: 206). The problem of distinguishing between play and non-play becomes evident when considering our ability to move in and out of a play and reflect upon it from the outside. The duality between being totally immersed in the game, while at the same time being able to reflect upon it from the outside (‘it is just a game’) is especially important when contemplating the ambiguity of describing activities as play [26].

James Carse [27], in turn, distinguishes finite from infinite games. The finite game, like Huizinga’s notion of play, is bounded and played for the purpose of winning. The infinite game, like life itself, is ongoing and is played for the purpose of continuing the play: “If the rules of the finite game are the contractual terms by which the players can agree who has won, the rules of the infinite game are the contractual terms by which the players agree to continue playing.” (p.9) In the infinite game, playing is a way of interacting. Finite games are about the outcomes (training to win and obtaining prestigious titles), while infinite games are about surprises as these generate new possibilities to reconsider the world, revealing a new beginning “like the grammar of a living language” (p. 9). As Carse aptly states: “To be prepared against surprise (as in the finite game) is to be trained, to be prepared for surprise is to be educated.” (p. 19) Whilst training regards the future to be finite (yielding a clear end result), “(e)ducation leads towards continuing self-discovery” (p. 19). In terms of our fields of concern, academia

and healthcare, we are interested in how these finite and infinite games play a role in dealing with evaluative metrics: how evaluative metrics produce norms for ‘good’ academic and healthcare work and encourage scholars and practitioners (and their managers alike) to reach their goals by playing the finite game. The infinite game, then, would be about changing the grammar of academic and healthcare work; prompting new and embodied routines of doing research and providing care.

The separation of games from non-games, or Carse’s distinction between finite games from infinite games, is crucial when discussing another aspect of our matters of play: gamification. In fact, the concept of gamification itself rests on this separation, with the application of game features in non-game contexts being a common trait [28]. Gamification aims to “transplant” some of the motivational qualities of games – especially points and scores, displaying performances and the desire to level up and win – into contexts that are not inherently leisure-focused or motivating in themselves [29, 30]. Raczkowski [29] discerns three features of games and gamification: (1) games as experimental techniques; (2) games as sources of flow; and (3) games as governed by points and high-scores. Games as experimental techniques refer to the more or less risk-free environment of a game in which certain tests but also training can be conducted in a less expensive way and without the fear of consequences beyond the game-world. Think for instance of a skills-lab in which a medical resident can practice a surgical procedure on a computer instead of a real patient [31]. Game as a source of flow addresses the ability of game features to transform daily routines in optimal (or optimizing) experiences by turning them into personal meaningful games. Lastly, games as governed by points and high-scores portrays the ‘magical’ aspect of gamification; the speaking to pleasure, convenience and personal entertainment [32]. Here, games simulate value through measurements that feed into circuits of reproduction, making performance visible and comparable, inserting an element of competition – typically through leaderboards. Pantzar and Shove [33] argue how such measurements feed into circuits of reproduction of everyday practice, linking micro-level performance to macro-level organization while simultaneously spanning past, present and future.

This quantified ‘points and high-scores’ feature of gamification is most elaborated on in the (STS) literature on gaming and the flourishing stream on personal analytics and self-tracking [20, 29, 34, 35]. Self-tracking has gained attention with the emergence of The Quantified Self movement [36]; people gathering quantitative data about themselves, using mobile apps and always-on gadgets, seeking to convert previously undetected bodily functions and behavioral clues (e.g. heartbeat, the number of steps taken) to traceable and perceptible information [34]. Self-tracking enables the making of “data doubles” [37]; ways of knowing and valuing the self in rather abstract and sliced ways, that can be reflected on and used for various purposes like self-optimization. Ruckenstein [38] points out how this theme of visibility – and, we like to add, ‘invisibility’, because by making some aspects visible, others are also (deliberately) ignored [38] – links personal analytics to modern notions of control and governmentality. With the aid of digital technology, optimization becomes not only possible but also desirable and an associated responsibility to act [39]. Through making unknown aspects of the body – or, in our case, professional performance – detectable and evaluable, we

(or, for that matter, ‘others’) can gain more control over life or professional ways of doing and knowing, reconfiguring existing professional identities.

Gamification and self-quantification can also have liberating effects, revealing a more ‘infinite’ aspect of gamification. Digital platforms like ResearchGate allow scholars to constitute and ‘better’ their scholarly reputation [20], for example by putting a project on ResearchGate and downloading and liking papers of co-authors – who are then enticed to ‘like’ the other in return. Others have pointed out researchers’ and particularly group leaders’ abilities to renegotiate values attached to evaluative metrics, for instance through negotiating how a specific journal is valued in a certain academic field and how publication achievements count in career mobility decisions – showing how evaluative metrics themselves become renegotiated, worked around and used by professionals [19, 40] and are embedded in wider practices of valuation.

In public administration and sociology, ‘gaming’ has also come to notify the ways in which practitioners and organizations ‘work around’ performance evaluation. For example, Espeland and Sauder define gaming as “a more or less cynical effort to manipulate the ranking data without addressing the underlying condition that is the purpose of the measure. It’s about managing appearances without really changing anything.” [41] Although we recognize that there might be cynical elements in playing evaluative games, we want to turn our attention to a type of play that is more generative in nature and is attuned to creating conditions for doing professional work. This is the reason why we link our notion of play to those of matters of concern (that is, the re-staging of representative ‘objects’ as lively matters of interest that must be engaged with, [22]) and matters of care (the relating to objects or ‘things’ by affectively engaging with them and exploring their potential to affect others, [23]). That is, we see play as the serious matter of caring for and working with the things that are of concern through the use of numbering practices.

In this paper we use the notions of play, games and gamification – hence resembling in the notion ‘matters of play’ – as analytical lenses to study how evaluative metrics produce professional practices and how these practices of (e)valuation are reshaped in return through the playful actions of professionals. Matters of play, we will argue, are relational practices that encompass concerns for professional practice (i.e. academia and healthcare), and the caring work this implies. This involves the play of both finite and infinite games, revealing the dispersed work and distributed effects of valuation practices and how these affect professional identities.

### **3 Research settings & Methods**

This paper is part of an ongoing collaboration between two research groups, one situated in the field of academic evaluation and the other in healthcare, both intrigued by the politics of evaluative metrics and its influence on professional work [6, 35, 42]. We draw on separate research projects that we bring together for the purpose of our analytical questions. We do not so much compare clear or ‘neat’ results between both fields (i.e. their similarities and differences) but rather embrace the messiness and hence lack of clarity in both fields in their dealings with evaluative metrics, mobilizing insights

from one field to elucidate and discuss findings in the other field, and the other way around [43].

In academia, we draw on a primarily interview-based research project in law faculties in the Netherlands, investigating how epistemic dynamics of scholarship are mediated through emerging evaluation schemes, in particular through the ‘micropolitics’ of indicator use [40]. In healthcare, we build on an ongoing research project in which we study how quantified evaluation practices (i.e. performance indicators, rankings) influence (the valuation of) medical and nursing work as well as healthcare organizations [4, 35].

## **4 Findings**

### **4.1 Academia: Politics of metrics and the politics of a discipline**

In the past three decades, academic research has been increasingly governed through evaluative metrics; i.e. numbers of (international) publications and citations, position in university rankings, and obtained (international) research grants at the individual, group and institute level. Evaluative practices (‘what counts?’) impact the design of research programs, the establishment of endowed chairs, and the finance of research projects. Doing well in those aspects, as an institution or research group, but also as an individual scholar, is crucial to building a good academic reputation – and subsequently gaining more rewards. In this section we turn to legal scholarship in the Netherlands, scrutinizing how finite and infinite games are played in this particular academic field. Law is a salient example as substantial parts of legal scholarship are domestically oriented and projects often focus on policy or national legal issues that may not be immediately interesting for a wider academic audience. This tension is particularly prominent in the case of non-English speaking countries like the Netherlands, where it has given rise to a complex politics of valuing publications and other forms of ‘output’ in assessment contexts.

As part of a comprehensive national evaluation system across all fields of research, Dutch law departments are subjected to an assessment exercise in a regular interval of six years. While criteria of ‘societal relevance’ have recently been strengthened, evaluation modalities continue to place significant emphasis on the publication output of departments. This creates a particular challenge for the epistemic culture of law: English-language, journal-based publications are weighed against the tradition of writing books in the national language and having a more practical orientation. The mismatch between the value ascribed to international publications and a strong professional, domestic orientation of the field is often seen to require an explicit disciplinary strategy [44] – a need to react to evaluative expectations of funders and policy actors by making a particular “move” in a larger game, namely the competition for prestige and resources with other academic fields. However, we encountered very different opinions among legal scholars about what kind of move this should be. On the one hand, there are those who advocate a ‘scientific turn’ of legal scholarship, i.e. a research style that seeks to generate relevance not just by focusing on the national legal system, but through its

methodologically and conceptually rigorous approach. Another position is to instead push for altering evaluative conventions in such a way as to confer greater appreciation for activities relevant to professional audiences – both on the level of the national evaluation exercise and the more implicit expectations of funding bodies towards publishing behavior.

These disciplinary arguments inter alia play out through a specific technical-administrative venue, namely diverging local practices of registering the collective departmental output that constitutes the basis for the regular evaluation exercises. In the course of the last years, all Dutch universities have adopted centralized digital information systems through which researchers are expected to register their publications and other activities. The different ways in which individual actors can access this information infrastructure affects their ability to enact professional routines as either finite or infinite games. To regular researchers, the need to register output often appears as an administrative chore that is squeezed into late working hours at the end of the month or semester. Seemingly mundane aspects such as user friendliness here directly affect the likelihood that certain activities are rendered visible. For example, when publications need to be manually entered rather than being automatically detected by the software – as is often the case for small-scale Dutch publication venues –, there is a chance that users simply forget to register them. Another aspect is that the use of a digital information system creates a certain barrier to what researchers consider ‘worthy’ of explicit registration in the first place. For many legal scholars, activities that would potentially qualify as societal engagement – such as public lectures, elective courses, consulting work for diverse public and private organizations – appear so naturally as part of their job profile that they often do not bother to register them in the first place.

On the other hand, a prominent role is reserved for the so-called program coordinators, i.e. scholars in group leader positions who are tasked with overseeing and signing off on the output data submitted by regular department members. Their administrative responsibility allows program coordinators to adopt a panoramic view on the totality of the activities of a research group. These collective activities here are much more easily perceived as a resource in an infinite, knowledge-political, game – should a group position itself as more societally relevant, as academically excellent, or in terms of some other specific institutional profile? A particular strategic opportunity for the program coordinators arises from the fuzzy distinction between academic and professional journals in the field of law. Often, it is unclear how to deal with particular publication genres (such as case annotations), and how to determine what audience(s) a journal is exactly catering to. Program coordinators promoting a ‘scientific turn’ tend to police the formal distinction between academic and professional journals particularly strictly, with the intention of ‘disciplining’ individual researchers to publish in rigorously reviewed, preferably international, outlets. The following quote describes a case where a program coordinator would reclassify publications submitted as ‘academic output’ by members of his department, but which in his opinion did not live up the criteria of rigorously peer-reviewed publications.

(...) when I joined this law faculty, legal scholarship here was more traditional than what I was used to [from working at another university]. ... lots of

annotations, lots of professional publications, resistance against a real scientific turn in our field. And I really had a lot of discussions about this. Specifically, the cause of the conflict was – and this used to happen a lot more in 2008 and 2009 than now, because now the hygiene has improved in this respect – so the conflict played out like this: I took them out [i.e. re-moved publications of program members from the list of peer-reviewed out-put] only for them to go upstairs to the fourth floor, where the administration was, and to say ‘no, this has to be reversed, I’m telling you, I want to see this reversed’. So things were reversed behind my back! But I always had the last laugh, because whenever the reports for the evaluation committee had to be prepared, I took them out again, even though I had to do double work. (Professor of administrative law, 21 April 2015)

Other program coordinators instead adopt the position that too strict a focus on purely academic audiences threatens to disrupt the current embedding of scholarly work carried out at university departments to the domestic legal system, thus under-mining a key function of the field in Dutch society. They tend to see the very fuzziness of the boundary between professional and academic publications as a virtue, as something that enables academics to be useful to diverse stakeholders across academia and legal practice. This position is nicely illustrated in the following quote by one of the program coordinators we interviewed. When it comes to double-checking whether publications of his research group are correctly classified as either scholarly or professional, this researcher generally adopts a laissez-faire approach. His reasoning is that a strict enforcement of the distinction is problematic and undesirable, given the heterogeneity of different intellectual approaches in legal scholarship. The only respect in which he makes proactive use of his administrative power is the way pro-gram members register professional output. He actively encourages members of his research program to highlight as many of their nonacademic activities as possible, thereby attempting to alter the hierarchy of value ascribed to different types of out-put.

My problem with the Metis-lists [the online information system used for registering output] is that people do not register a lot of validation activities. They have always registered their publications, but not their contract work. Not the 20 times per year that they teach a course. They also don’t register it when they present a paper in a congress or participate in an expert meeting. In my own case, 10 out of 100 items are publications. 90 are other things that can be registered through Metis and that should be in there, if you want to show the relevance of your work. (Professor of private law, 05 May 2015)

The above examples show how finite and infinite games are intertwined. On the one hand, the need to produce publications in particular types of outlets constitutes a finite game from the perspective of individual researchers, with implications for their career development in institutions and the competition for funding. Program coordinators leverage these finite games in diverse ways in their attempts to make a move in a larger, infinite game, namely the perpetual struggle between different intellectual positions in the disciplinary research culture of law. In turn, these higher-level moves have implications for the perception of the field by funders and policy-makers.

### **Counting as caring for the group**

However, not all program coordinators use their specific administrative power with the intention of making a particular strategic move in the larger debate about the orientation of legal scholarship in the Netherlands. Instead, some of them simply opt to continue playing the evaluation game in its current form, but in such a way as to try and maximize their success in both evaluation and the acquisition of funds. For example, many law departments are particularly dependent on income from contract work, including public bodies such as national ministries and institutions of the European Union. Such work often puts significant strain on researchers to combine commercial activity necessary to maintain their institutional revenue streams with re-search proper. A particular complicating factor are contractual obligations that frequently prevent reusing empirical material generated for commissioned projects as a basis for publishing academic articles. One of the program leaders described a particular way of quantifying output that makes it easier for his researchers to reconcile the twin mission of conducting both ‘excellent’ academic research and contract work. More specifically, he is inclined to count particularly thorough pieces of contract work as scholarly output, even though they would formally have to be classified as professional:

The reason why I reclassify more than others in this department within this university, or within this law school, is that the research groups like we are, have a much higher financial target, because we do less education (...). So, two thirds of the budget I need to earn outside (...). We do a lot of work also for the European Union. And that also is published in books or in articles, but sometimes in reports. And I think it’s unfair if I say, this report is just as scholarly as this book, but I cannot calculate it because it has not been published in a book or because the Commission prohibits to publish again. Then I cannot give credits to something that is a research project of two years with the exactly the same work as if it was not externally funded. So, then I give credits for those outputs. (Professor of international law, 30 June 2015)

Classifying research output and subsequently counting their worth here is part of the caring craft work [45] that group leaders conduct to enable their research groups to flourish in the academic world. Group leaders like the professor quoted above seek to mold their research context and make it fit into the evaluative metrics system. This can be seen as a finite game: the game is strategically prepared and played in such a way as to make financially attractive yet unpublishable research count nevertheless. But while such practices are not meant as a stance in the broader debate about a ‘scientific turn’, they nevertheless have subtle implications for the disciplinary re-search of law. Through making it easier to reconcile principally different types of activity – contract work for legal stakeholders and academic research proper –, the ‘grammar’ of law research at particular departments is being shaped in the context of longer-term research and societal goals. Again, infinite games thus comprise finite ones – being relevant to societal actors in the Netherlands or an EU level entails stretching the rules of the finite game.

## 4.2 Healthcare: Finite and infinite games to sparking good care and clinical reputations

### Caring for reputation

Also in healthcare, evaluative metrics have gained great importance in recent years. Performance indicators, like the number of pressure ulcers on a nursing ward, the percentage of readmissions after hip surgery and the measures on pain scores among oncology patients (to name a few), are increasingly considered important instruments to assess (and compare) quality of care. In the Netherlands, performance indicators are used among a wide variety of healthcare actors (insurers, regulatory authorities, patient organizations, professional associations) to define and improve quality of care. Health insurers, for example, use performance indicators to contract preferred providers – for instance those hospitals that meet the quality requirements of the breast cancer associations (amongst others, percentage of remaining cancer tissue after surgical treatment, presence of a specialized nurse, waiting time – also illustrating the diverse range of quality aspects that are assembled in the indicators). Furthermore, based on the scores on performance indicators, league tables are composed displaying the best hospitals. Such league tables are widely discussed, both in the media and among healthcare providers. And although practitioners and managers alike are highly ambivalent towards the hospital rankings – as one hospital manager exclaimed: “the one year you can be on top, and the next you may drop to a much lower level without actually changing your care policies” –, it has also enabled healthcare managers to come to grips with governing care provision, moving it out of the direct control of the medical profession [4]. The displaying of quality of care, and with that the hospital’s reputation, has turned (measured) health care quality into a shared concern and strategy:

Managers and doctors are hotly discussing whether lowering the age from 70 to 60 [for screening on delirium] makes sense, and whether screening for delirium is a useful indicator [with regard to investment and return for patient safety] after all. The executive ends the discussion and argues: “We have to put a stop to the discussion. We lose out in important points in the [ranking] due to the delirium case. This simply has to happen.” (Observation, Hospital C management team, November 15, 2012)

This ‘has to happen’ argument has become more common in medical practice; evaluative metrics have changed the rules of the game, rendering healthcare managers, insurers and government authorities more central players [4]. Indicators enable hospitals to gain a good reputation (being seen as an ‘outstanding hospital’) and through that gain rewards, for instance by being appointed as a preferred provider. Furthermore, a good (quality) reputation has a liberating effect as having good scores and obtaining a high-ranking position provides confidence to regulatory authorities (i.e. the inspectorate, health insurers) that hospital executives and clinicians are in control of healthcare quality, leaving them with more room for maneuver – and thus less external control. Such hospitals are also inclined to attract high-qualified practitioners; being a preferred provider or obtaining a prestigious quality certificate or research license attracts ambitious physician and nurses. Hence, the finite game of obtaining a high ranked position yields

infinite elements as well, as the use of evaluative metrics transforms the grammar of hospital governance; good quality becomes a shared game.

### **Bending indicator thinking**

Hospitals have installed dashboards to monitor the scores on performance indicators and compare nursing wards on their achievements (e.g. to what extent they comply with the protocol to measure and jot down the score on fall risk among elderly patients three times a day). During our observations, we noticed hospitals executives carrying around print-outs of the dashboard scores when visiting clinical groups and nurse managers 'to insist on the importance of indicator use'. Furthermore, we observed a staff meeting of quality managers discussing how they could best visualize the scores on performance indicators (translating scores into diagrams and flow charts) in order to induce an element of competition between wards to encourage them to do (that is, score) better. As one of the informants explained:

It is demotivating if wards remain in the red area [representing the lower scores] for longer periods of time. In the new reports, we don't make it all red and orange but focus on the green; how you can work towards the green zone. This is the first month we do it like this, and now we wonder: does this motivate them enough? Because now people may no longer notice they are actually in the red zone. So maybe we will add an orange area again. (Hospital quality staff, 19 September 2016)

This excerpt reveals the 'points and high-scores' feature of gamification; installing elements of competition in order to 'do better' (fulfilling the indicator requirements). It also shows the messy reality of indicators and the tinkering involved in making sense of indicator scores; for staff quality indicators have become strategic attempts to standardize professional work. Professionals, however, may ignore or work around these quality policies. One of the nursing ward managers was quite skeptical about indicator use. He demonstrated the dashboard of his ward on the computer screen to us, showing how he could 'real time' monitor whether the nurses conducted the measurements according to the protocol. Yet, to him it was only important that the nurses conducted those measurement that mattered to the clinical and social condition of the patient, and that they did not waste their time on what he thought were bureaucratic activities:

I want them to use their clinical gaze; what are the risks their patients face? If a patient runs the risk of getting a delirium, I want them to measure their patients three times a day [using the delirium score list], no doubt. And if they don't do these measurement, I'll let them know. But if a patient walks around without showing any signs of pain, it's ridiculous to do the pain scores. I will fill them out for them. (Nurse ward manager, 15 November 2017)

This nurse ward manager played the indicator game by embracing the indicators that help to improve care ('signaling an emerging delirium in order to take preventive measures in time') while only cosmetically complying with those that he thought to be redundant as they do not contribute to the quality of caring work; the manager thought up (and registered) those scores himself in order to not fall behind in overall scores and end up low on the hospital's internal (and, in the end, national) ranking. Whilst the

quality manager, expectantly, was quite cynical about this manager ‘ignoring’ important quality improvement work, an anesthesiologist involved in fabricating the national pain indicator reacted with a smile when sharing our anecdote: “They at least consider the possibility of pain, that is much more important than actually doing the score; you have to really think about it.” (anesthesiologist, 5 October 2017)

### **Mobilizing care improvement**

The micropolitics of indicator-use [40] further play out in mobilizing (new) care routines. Indicators actively produce care practices as they set the norm for ‘good care’. The Dutch Healthcare Inspectorate plays a central role herein. It employs indicators as strategic instruments to evoke other ways of organizing care. A striking example is the norm of involving a geriatric physician in the treatment of patients suffering from hip fracture:

It took us years to get this done...and we’re still underway. Hip fracture was one of our first indicators and has changed over the years. We became increasingly aware that hip fracture primarily is a social problem. It is not so much about fixing a fracture, that’s easy, but about the social problems frail elderly experience at home. This should be taken care of. So, we mobilized the geriatric physicians. They were a bit hesitant to play a bigger role [in the treatment of hip fracture], but we managed to move them into a central position. Now they are part of the [indicator’s] definition. [Interview inspector, 7 March 2018]

This excerpt demonstrates how the inspectorate has played a strategic game by moving the geriatric physician in a central caring role in case of hip fracture. Through rendering the geriatric physician part of the indicator definition, the involvement (and, in some cases, employment) of a geriatric physician was set as a norm for good hospital care. This example reveals the infinite game that is played to improving care. The old hip fracture indicator is reframed to adjust vested hospital practices to the increasing problem of a growing group of elderly patients admitted to the hospital because of urgent clinical problems that actually mirror troubled situations at home where elderly people suffer from loneliness and lack of care. Through adjusting the hip fracture indicator, positioning it as a finite game (hospitals compete on scoring levels for high scores on the Inspectorate’s indicators) the inspectorate sought to mobilize social care for elderly patients in the hospital and, through that, at home. This playing hence comprises recognition and anticipation of the performative effects of the new practices that come with metrics – rendering them a matter of concern and care.

## **5 Discussion**

Evaluative metrics have entered academia and healthcare as metering and disciplining techniques to improving and accounting for professional performance, entering a point in which they are increasingly considered ‘necessary for survival’ – turning into an evaluative monster that begs for attention and ever more data, scores, benchmarks and resources. As Latour [46] remarks, however, the monster of Frankenstein only became

a monster after it was left on its own by its creator. In similar fashion, today's 'metric monsters' are not monstrous from birth, and subsequently their 'evil' is not in their design, but in how they come to be used and cared for – or not. In fact, many professionals in health care and academia do try to find a way of living with their monsters, and caring for them becomes an integral part of the daily work – also, or maybe particularly because, they can be beneficial to them. This form of care can consist in putting the metric monster to work when needed (when a patient is in risk of delirium), and playfully resisting (rather than ignoring) it when its services are unwarranted. As we know, monsters have great strength, and when used strategically – as with the example of the hip fracture indicator – metrics can, when cared for, be utilized to change established practices which would be hard to transform without the magical power of indicators and scoreboards.

The analytical distinction between finite and infinite games has allowed us to open up the ways evaluative metrics are enacted in daily professional practice, but also to understand how particular metrics are often leveraged in conflictual dynamics. A key point here is the reflexivity that underpins the use of metrics. Actors not only 'do' measurements but also reconsider and negotiate them, using metrics to reach particular goals but also playfully ignoring them when they are not considered useful. As a finite game, individuals will often see metrics as a force set in motion by remote superiors or a faceless policy apparatus – something that must be obeyed and can only be made beneficial by using it for one's own purposes (e.g. building up a scholarly CV according to metrics criteria). In finite games, professional practices in academia and healthcare become contests that have winners and losers, stimulate game players to live up to the rules of the game, and at the same time change epistemic and caring work by reframing them in the language of evaluative metrics uses. In the infinite game, however, evaluative metrics do not so much appear as an obtuse monolithic force, but rather are leveraged in a reflexive manner – think about the pain indicator forcing clinicians to at least consider the possibility of pain when treating a patient; or, turning to the academic context, categorizing scientific output in certain ways to allow for different relations between academic and practical work. Infinite games are played not to follow rules, but for continuing the game itself. In doing so, they create new possibilities and vocabularies for professional work. The very effects of such moves (e.g. a new practice of counting of publications) may in turn however be perceived as monstrous by others – elucidating how finite and infinite games in practice play together, as well as create tensions.

This leaves us with some compelling concerns, hopes, and desires for further research. For one, while playing together may at first sound peaceful, Huizinga actually reminds us that playing can involve cruel behavior and suffering as well. Tensions may arise not only when players compete with each other in well-defined finite games, but also when the more strategic, infinite, games restructure practices in unforeseen ways. Think for example of a situation where researchers embrace a form of evaluation that may be useful to themselves or their institution in the short run, but that also creates more undesirable longer-term consequences for their field as it heightens (publication, administration) expectations for all. A problem here is that the reflexivity that underpins particular attempts to strategically leverage evaluation practices is always bounded. Individual actors may act in ways they believe are in the best interest of their professional

group or field, but in reality, the effects of their moves may cause unpredictable reactions and counter-moves by other players.

A simultaneously analytical and political problem – to be pursued in further work – is that strategic agency in evaluation games is usually unequally distributed. Some actors thus are more capable of making moves in infinite games, whereas others are primarily stuck in reactive positions [8]. For example, nurses may experience more pressure to fulfill scoring requirements than doctors who are in a better (i.e. more powerful) position to legitimately question evaluative metrics or bend metric rules in ways that allow them to pursue their professional goals. Similarly, researchers in important administrative positions at universities are clearly more powerful to locally interpret evaluative practices than their colleagues. A comparable dynamic is likely to play out on an institutional level. Elite universities are more likely to shape evaluative conventions (either by proposing their own research practices as an evaluative model, or by refusing to follow the “latest trend” in research and evaluation), whereas less well-reputed universities are more dependent on short-term approval and funding by policy [47]. As Espeland and Sauder [7: 188f] argue, powerful actors have greater possibility to make use of inherent ambiguities involved in evaluation and measurement, and thus ‘change the rules of the game’.

We also have hopes, however, that tie into our research ideas. Play, games and gamification are useful concepts that point out the reflexivity and relational aspects of evaluative metrics. Metrics not only enforce particular forms of accountability, but also render visible new and experimental forms of performance governance. Recognition of those performative effects of metrics also turns them into matters of concern, care – and play.

## References

1. Kelly, A. and R. Burrows, Measuring the value of sociology? Some notes on performative metricization in the contemporary academy. *The Sociological Review*, 2012. 59(2): p. 130-150.
2. Rushforth, A.D. and S. de Rijcke, Accounting for impact? The journal impact factor and the making of biomedical research in the Netherlands. *Minerva*, 2015. 53(2): p. 117-139.
3. Essén, A. and M. Sauder, The evolution of weak standards: the case of the Swedish rheumatology quality registry. *Sociology of Health & Illness*, 2016. 39(4): p. 513-531.
4. Wallenburg, I., J. Quartz, and R. Bal, Making hospitals governable: performativity and institutional work in ranking practices. *Administration & Society*, 2016.
5. Burrows, R., Living with the h-index? Metric assemblages in the contemporary academy. *The Sociological Review*, 2012. 60(2).
6. de Rijcke, S., et al., Comparing comparisons. On rankings and accounting in hospitals and universities, in *Practising Comparison: Logics, Relations, Collaborations*, M. Press, Editor. 2016, Mattering Press: Manchester. p. 251-280.
7. Espeland, W.N. and M. Sauder, Rankings and reactivity: how public measures recreate social worlds. *American Journal of Sociology*, 2007. 113(1): p. 1-40.
8. Fochler, M. and S. de Rijcke, Implicated in the indicator game? An experimental debate. *Engaging Science, Technology, and Society*, 2017. 3: p. 21-40.

9. Hazelkorn, E., *Rankings and the reshaping of higher education: The battle for world-class excellence*. 2011: Palgrave Macmillan Publishers.
10. Fochler, M., U. Felt, and R. Müller, *Unsustainable growth, hyper-competition, and worth in life science research: Narrowing evaluative research repertoires in doctoral and postdoctoral scientists' work and lives*. *Minerva*, 2016. 54: p. 175-200.
11. Hammarfelt, B. and S. de Rijcke, *Accountability in context: effects of research evaluation systems on publication practices, disciplinary norms, and individual working routines in the faculty of Arts at Uppsala University*. *Research Evaluation*, 2014: p. 1-15.
12. Müller, R., *Crafting a career in STS: meaning making, assessment, and interdisciplinary engagement*. *Engaging Science, Technology, and Society*, 2017. 3: p. 84-91.
13. Levay, C. and C. Waks, *Professions and the pursuit of transparency in healthcare: two cases of soft autonomy*. *Organization studies*, 2009. 30(5): p. 509-527.
14. Bozeman, B. and D.M. Anderson, *Public policy and the origins of bureaucratic red tape: implications of the Stanford yacht scandal*. *Administration & Society*, 2016. 48(6): p. 736-759.
15. Hammarfelt, B., S. de Rijcke, and P. Wouters, *From eminent men to excellent universities: university rankings as calculative devices*. *Minerva*, 2017.
16. Wallenburg, I., et al., *Onderzoek naar risicoselectie met de basisset kwaliteitsindicatoren ziekenhuizen: op weg naar verantwoorde keuzes*. 2018, Amsterdam Public Health: Amsterdam.
17. Berg, M., et al., *Feasibility first: developing public performance indicators on patient safety and clinical effectiveness for Dutch hospitals*. *Health Policy*, 2005. 75: p. 59-73.
18. Bonde, M., C. Bossen, and P. Danholt, *Translating value-based healthcare: an experiment into healthcare governance and dialogical accountability*. *Sociology of Health & Illness*, 2018.
19. Bal, R., *Playing the indicator game: reflections on strategies to position an STS group in a multi-disciplinary environment*. *Engaging Science, Technology, and Society*, 2017. 3: p. 41-52.
20. Hammarfelt, B., A.D. Rushforth, and S. de Rijcke, *Quantified academic selves. The gamification of research through social networking services*. *Research Information*, 2016. 21(2).
21. Bevan, G. and C. Hood, *What's measured is what matters: targets and gaming in the English public health care system*. *Public Administration*, 2006. 84(3): p. 517-538.
22. Latour, B., *Why has critique run out of steam? From matters of fact to matters of concern*. *Critical Inquiry*, 2004. 30(2): p. 225-248.
23. Puig de la Bellacasa, M., *Matters of care in technoscience: assembling neglected things*. *Social Studies of Science*, 2011. 41(1): p. 85-106.
24. Huizinga, J., *Homo Ludens: a study of the play element in culture*. 1955, Boston MA: Beacon Press.
25. Rodriguez, H., *The playful and the serious: an approximation to Huizinga's Homo Ludens*. *Game Studies*, 2006. 6(1).
26. Raessens, J., *Making points the point: towards a history of ideas of gamification*, in *Rethinking Gamification*, M. Fuchs, et al., Editors. 2014, Meson Press: Lüneburg, Germany.
27. Carse, J.P., *Finite and infinite games: a vision of life as play and possibility*. 1986, New York: The Free Press.
28. Raczkowski, F., *Making points the point: towards a history of ideas of gamification*, in *Rethinking Gamification*, M. Fuchs, et al., Editors. 2014, Meson Press: Lüneburg, Germany. p. 141-160.

29. Ruckenstein, M. and M. Pantzar, Datafied life: techno-anthropology as a site for exploration and experimentation. *Techné: Research in Philosophy and Technology*, 2015. 19(2): p. 193-212.
30. Deterding, S., et al., Designing gamification: creating gameful and playful experiences, in *CHI 2013: Changing Perspectives*. 2013: Paris, France.
31. Johnson, E.S., Out of My Viewfinder, Yet in the Picture: Seeing the Hospital in Medical Simulations. *Science, Technology & Human Values*, 2008. 33: p. 53-76.
32. Whitson, J., Gaming the quantified self. *Surveillance & Society*, 2013. 11(1/2): p. 163-176.
33. Pantzar, M. and E. Shove, Metering everyday life, in 17th Annual SASE Meeting. 2005: Budapest.
34. Lupton, D., *The quantified self: a sociology of self-tracking*. 2016, Cambridge: Polity.
35. Wallenburg, I. and R. Bal, The gaming doctor/nurse: How practices of datafication and gamification 'redo' care. *Health Informatics Journal*, 2018: p. doi 10.1177/1460458218796608.
36. Wolf, G., The data-driven life, in *The New York Times Magazine*. 2010.
37. Ruckenstein, M., Visualized and interacted life: personal analytics and engagements with data doubles. *Societies*, 2014. 4(1): p. 68-84.
38. Pinto, M.F., Tensions in the agnotology: normativity in the studies of commercially driven ignorance. *Social Studies of Science*, 2015. 45(2): p. 294-315.
39. Viseu, A. and L. Suchman, *Wearable augmentations: Imaginaries of the informed body, in Technologized images, technologized bodies: anthropological approaches to new politics of vision*. 2010, Berghahn Books: Oxford, New York.
40. Kaltenbrunner, W. and S. de Rijcke, Quantifying 'output' for evaluation: administrative knowledge politics and changing epistemic cultures in Dutch law faculties. *Science & Public Policy*, 2016. 44(2): p. 284-293.
41. Espeland, W.M. and M. Sauder, *Engines of anxiety: academic rankings, reputation, and accountability*. 2016, New York: Russell Sage Foundation.
42. Hammarfelt, B., et al., *Advancing to the next level: the quantified self and the gamification of academic research through social networks*. 2017.
43. Deville, J., M. Guggenheim, and Z. Hrdličková, eds. *Practising Comparison: Logics, Relations, Collaborations*. 2016, Mattering Press: Manchester.
44. Stolker, C., *Legal Journals: in Pursuit of a More Scientific Approach*. *European Journal of Legal Education*, 2005. 2: p. 77-94.
45. Davies, S.R. and M. Horst, Crafting the group: care in research management. *Social Studies of Science*, 2015. 45(3): p. 371-393.
46. Latour, B., Love your monsters: why we must care for our technologies as we do for our children. *The Breakthrough Journal*, 2012. 2: p. 21-28.
47. Paradeise, C. and J.C. Thoenig, Academic institutions in search of quality: local orders and global standards. *Organization Studies*, 2013. 34(2): p. 189-218.