Gaming the system: building an online management game to spread and gather insights into the dynamics of performance management systems

Scott Douglas, Christopher Hood, Tom Overmans & Floor Scheepers

Published online: 04 Feb 2019.

Submit your article to this journal ⬚

Article views: 909

View related articles ⬚

View Crossmark data ⬚

Routledge
Taylor & Francis Group

Gaming the system: building an online management game to spread and gather insights into the dynamics of performance management systems[*]

Scott Douglas[a], Christopher Hood[b], Tom Overmans[a] and Floor Scheepers[c]

[a]Utrecht School of Governance, Utrecht University, Utrecht, The Netherlands; [b]All Souls College, Oxford University, Oxford, UK; [c]Utrecht Medical Center, Utrecht University, Utrecht, The Netherlands

ABSTRACT
Extensive research has produced many insights into the dynamics of performance management systems. Spreading these complex insights among students and practitioners can be a daunting task. Gathering new insights can be equally challenging. This article introduces a novel tool for teaching and researching performance management, reporting on the design and first use of a free online management game. Players take the role of a hospital manager trying to satisfy multiple stakeholders through applying different performance management instruments. While students learn about the complexities of performance management, researchers gather data about the pathways individuals pursue while navigating performance management systems.

Introduction

The performance management literature has produced an expansive range of insights in the way public-sector employees, managers, and stakeholder shape and respond to performance management and measurement (for an overview see, for example, Gerrish 2016, 48–66; Van Dooren, Bouckaert, and Halligan 2015). This has expanded the focus from merely mapping performance indicators to analysing the dynamics of the whole system of instruments, routines, and mechanisms (Moynihan and Kroll 2016). Spreading these complex insights to students and practitioners can then be a daunting task. How to condense the findings from decades of research? How to transmit nuanced observations about the unpredictable effects of performance management instruments? Gathering new insights into performance management systems can be equally challenging. How to observe the effect of performance management information on decisions? How to examine the trade-offs actors make between conflicting demands such as efficiency, legitimacy, and quality?

CONTACT Scott Douglas ✉ s.c.douglas@uu.nl

Insights into performance management are currently spread and gathered through multiple methods. Detailed case studies give researchers and students a rich understanding of the dynamics of performance management in practice (e.g. Nomm & Randma-Liiv 2012). Surveys and vignette experiments shed light on the ways different public actors define performance, use performance information, and apply management instruments (e.g. Lewandowski 2018). Meta-analyses and effect studies reveal the benign or perverse impacts of performance management instruments on organizational outcomes (e.g. Bevan and Hood 2006; Gerrish 2016). The combination of these different methods makes for a strong teaching and research toolkit, but there are some lacunas as well. Lectures or articles do not always allow for engagement and reflection on part of the students (Ten Cate et al. 2004, 219–228). Good case studies require extensive time commitment for both fieldwork and classroom discussion. Focused surveys and experiments may lose sight of some of the richer dynamics of performance management systems.

This article explores to what extent an online management game could form an additional tool for teaching and researching performance management. Educational games provide an extra teaching method through individual immersion and experience (Connolly et al. 2012; De Smale et al. 2015), while also providing an additional way for gathering data by reaching bigger audiences while tracking individual choices (Mayer et al. 2014). This reflection is based on our experiences building a free online game placing players in the management seat of a hospital department. Players have to balance the competing demands of doctors, patients, and the chief financial officer. They get to experiment with different types of instruments (targets for patients treated per hour, benchmarking patients complaints per doctor, etc.) and receive information on how these interventions affect outputs, staff morale, patient satisfaction, and senior management support (see Figure 1 for the player dashboard). If they do not improve the performance of the department within four quarters, they will be fired.
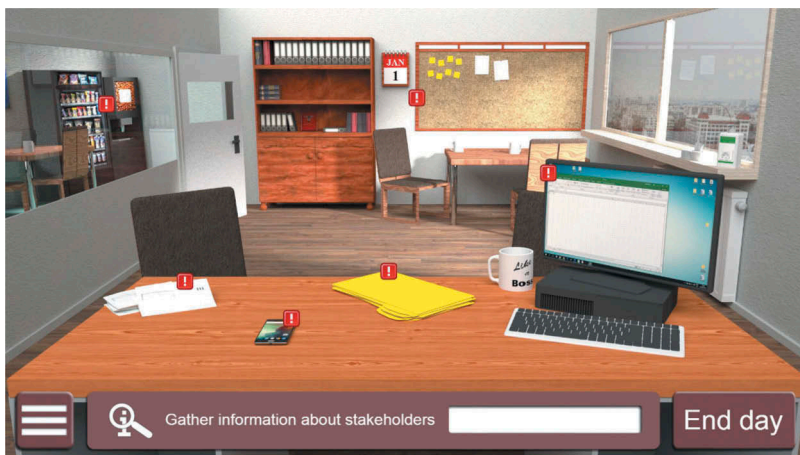
Figure 1. Player dashboard.
The home screen of the player gives access to different information sources, such as the management report on key indicators, a phone for social media reports from patients, a quarterly letter from the CFO, and a coffee machine to chat to medical staff. The computer can be used to introduce new performance management instruments, such as setting goals for different parameters of patient treatment and patient administration.

The main goal of the game is not to 'solve' the performance management puzzle, but for students to explore the choices and dilemma's involved and for researchers to gain insights into the choices and pathways players pursue. This article describes the learning goals selected from the rich performance literature and the difficult translation of these nuanced insights to the crisp algorithms required for the game design. The article also shows how the game data can be used to visualize and track the pathways of different players as they review information sources, set performance instruments, and adjust their parameters over several rounds. Serious games can be a potentially valuable tool for teaching and research. Games do add a fresh perspective, which allows for a novel exploration of individuals choices and pathways by students and researchers alike. However, as the crisp logic of game design sometimes struggles to capture the ambiguity of performance management, the use of different methods remains vital.

Selecting key insights and methods

Selecting key insights to spread

The aim was to create a game which could be played and discussed within in an hour so that public administration teachers could easily integrate the game into their already crowded lesson plans. At the same time, the game should avoid advocating a simplistic 'one best way' approach to performance management. The game therefore had to focus on a subset of the many relevant issues in performance management. The idea is that the game offers an introduction to some relevant topics within performance management, where the teacher can leverage the post-game debrief and other methods to raise further issues. The three issues to be addressed by the game are the role of (1) diversity, (2) complexity, and (3) contingency in performance management systems. This selection excluded such important topics as the challenge of managing performance within network settings (Emerson & Nabatchi, 2015) or the proactive management of the stakeholder environments (Torenvlied et al. 2013, 251–272).

Diversity

Since its introduction in the 1980s, performance management has proven to be so widespread and so enduring that it can no longer 'be regarded solely as the lynchpin in the New Public Management (NPM) movement […] It has gone much wider than that and has been used, in a variety of ways, by a wide range of approaches to reform and improvement' (Pollitt 2018, 168). Performance metrics often arise from a wish from the principal setting policies to monitor the agent delivering those policies (see for example Verhoest and Wynen 2016, 1–31). Simple steering – in the form of incremental adjustment based on simple commands in the form of marching orders of the day and direct observation in the form of 'management by walking around' – is increasingly rare in the public sector.

Metrics are found in organizational arrangements emphasizing contracts and targets (Van Dooren, Bouckaert, Halligan, 2015), but also in more fluid performance regimes favouring collaboration and public value creation (Talbot 2008, 1569–1591; Moore 2013). As a consequence, performance metrics necessarily mesh with the wider variety of the public domain, involving multiple actors, clashing priorities, and many potential strategies. Students and practitioners should gain an appreciation of performance management through metrics in all its diverse shapes and forms.

Complexity

The application of performance metrics, while sometimes presented by advocates as technical and even common sense in nature, often involves unintended, unanticipated or counter-intuitive outcomes (Bevan and Hood 2006, 517–538; Kerpershoek, Groenleer, and de Bruijn 2016, 417–436). Getting to grips with this unpredictability – such as why performance rankings seem to resonate in some times and places but not others or why pay-on-performance schemes intended to motivate public employees frequently have the opposite effect – requires careful analysis of the subtle ways that formal institutional arrangements can interact with the informal aspects of public service provision (Van Dooren, Bouckaert, and Halligan 2015). This can result in not only unintended effects, but also perverse effects such as output distortion and threshold effects (Bevan and Hood 2006). The emerging insights go well beyond simple precepts such as 'what gets measured gets managed'. The insights of research into the unintended effects of performance metrics mean that students and practitioners need to appreciate this complexity of performance management as well.

Contingency

Although the debates over metrics in politics and public management often diverge into pro- and anti-quantification positions, such polarization can obscure more fruitful questions about what metrics work when (Hood 2012). The contingency perspective on performance indicators is a recurring approach in both public and private management, and in the former case attempts to identify how we can combine legitimacy, effectiveness, efficiency, and adaptation within highly variable and often changing contexts of public organizations (Calciolari, Prenestini, and Lega 2018, 1–23; Meier et al. 2015, 130–150). The focus on these underlying questions requires moving beyond easy dismissal or uncritical endorsement of performance metrics. The aim of the game is to engage students and practitioners in a fruitful exploration of the different contexts in which performance management can be applied in different forms, turning students and practitioners into active explorers rather than simple exponents of pro- or anti-metrics ideologies.

Selecting key insights to gather

The game could also be a tool for gathering fresh insights into the ways individuals navigate performance management systems (cf. Mayer et al. 2014). Leveraging the ability of the game engine to anonymously track the clicks and choices of players, this game could focus on the actions individuals take in the face of all this diversity, complexity, and contingency. As any computer game is ultimately a controlled environment, (Narayanasamy et al. 2006), the game may be unable to address research questions requiring greater participant freedom, such as asking players to define performance for themselves or find a radical alternative for the rationalistic approach central to performance management. We therefore wanted to enable the game to specifically document (1) performance information use, (2) performance instruments selection, and (3) response to different stakeholder interests. This in-game data could be matched with data provided by the players before the game starts about their age, education, professional background, and attitudes to public service.

Using performance information

Although public organizations have begun to generate large amounts of performance information, there is evidence that this information is not always well used (Moynihan

and Kroll 2016, 314–323; Angiola and Bianchi 2015, 517–542). The use of performance information seems to be contingent on individual characteristics and organizational routines (Moynihan and Kroll 2016). The game could offer the players a menu of information sources (e.g. a management report, survey among staff, etc.) and then track what they choose to consult or ignore. Together with the personal background information provided by the player before the game, the game could provide insights in the ways personal characteristics and organizational context shape the use of performance information.

Selecting performance instruments

As discussed, there is a rich diversity of performance instruments available. Although sometimes presented as a merely technical design question, 'design decisions are political and managerial decisions as well as technical ones, and they certainly have political and managerial consequences' (Pollitt 2018, 169). What instruments are applied may be dictated by the policy field and the wishes of the principal, but could also be shaped by the personal preferences of managers or national characteristics (Bouckaert and Halligan 2006). The game could offer players different types of instruments (e.g. targets versus rankings versus information sharing) and multiple implementation styles (e.g. top-down versus bottom-up target setting) and see what different patterns emerge in these choices. Again, the data provided by the players at the beginning about their background and preferences could be used to better understand these choices. Moreover, the game could track whether players adapt their instrument choice over multiple rounds to optimize results, providing insights not only in the way people try to measure performance but how they try to manage it (Gerrish 2016).

Balancing stakeholder demands

A key driver of the complexity of performance management in the public sector is the need to cater to different stakeholders and their respective needs (Speklé and Verbeeten 2014, 131–146). The same indicator 'is frequently assessed differently by different stakeholders' (Pollitt 2018, 169). In the case of a hospital, patients, medical staff, and financiers may give different priorities to good service, professional freedom, and efficiency. These values could be framed as clear trade-offs, but also a different forms of logics about what a public organization should do and how it should function (Kerpershoek, Groenleer, and de Bruijn 2016, 417–436). For example, Noordegraaf (2015) distinguished between a professional, political, and performance logic which stakeholders could use. Assuming a clear clash between these logics, doctors will not only place a higher value on care quality than a hospital CFO, but also believe this best pursued through bottom-up, staff-led initiatives rather than top-down managerial interventions. These logics may not always be at such odds with each other. Logics can indeed clash, subvert each other, or be traded off against each other, but can also applied in parallel in different situations or be intertwined into a new hybrid logic (Skelcher and Smith 2015). Again, we can use the pre-game provided personal data and the in-game choices to see what different patterns emerge in how players navigate different stakeholder demands.

Mobilizing the potential of computer simulation games

Serious games could be used to spread and gather insights into the dynamics of performance management systems. From a student perspective, the need to appreciate the diverse, complex, and context-bound nature of performance management

requires a combination of cognition (what to learn or skill), affection (why learn or will), and metacognition (how to learn or metaskill) (Connolly et al. 2012). *What* to learn is probably the easiest aspect since the knowledge about performance metrics can readily found in books or papers, although the large size of the research literature means that teachers inevitably have to select what findings are important to transmit. *Why* it is necessary to understand metrics may be harder convey to certain audiences, especially professionals such as doctors or judges who have acquired management tasks on top of their professional expertise but possess little interest in management as such. *How* to learn requires more individual reflection and coaching than can usually be offered in classic lecture hall and classroom settings, necessitating the use of other teaching tools.

Serious games and simulations can help to address the cognition, affection, and metacognition (Ten Cate et al. 2004, 219–228). Games such as *FlightGear Flight Simulation* or the *REALGAME Management Simulation* engage players personally by allowing them to individually experience the positive and negative consequences of their own decisions, utilizing intrinsic rather than external motivation techniques (Palmunen et al., 2013, Kusurkar, Croiset, and Ten Cate 2011, 978–982). Especially for the upcoming generation of students and practitioners, games provide a recognizable and convenient way of learning (Overmans et al. 2017). Games are fun, interactive, and can be played when and wherever you want. Importantly, however, while serious games can be powerful teaching tools, they ultimately rely (1) on the role and involvement of the teacher, (2) on the specificity of the game in relation to the learning objectives, and (3) on the quality of the offline debriefs with the teacher to achieve lasting learning outcomes (De Smale et al. 2015; see also Crookall 2010). Any game about performance management would have to work with both the strengths and limitations of this teaching method.

Serious games can also be used for research purposes. The most famous example is the use of the FoldIt game developed by the University of Washington to solve a puzzle about protein folding. By presenting the mathematical challenges to the general public as an online game, the research team generated solutions faster than an algorithm could have done (Cooper et al. 2010). Games have been used for research in psychology, ethics, and business studies (Calvillo-Gámez, Gow, and Cairns 2011), as they help to observe human behaviours (Vasilou & Caims, 2011) and the response to different scenarios (Smith and Trenholme 2009). However, serious games face several limitations in a research context. Interactions between online players tend to downplay the impact of emotions occurring when people interact in real life (Laird 2002). Participants with game experience at home tend to act differently in serious games than people new to computer games (Smith and Trenholme 2009). Equipping the virtual/artificial intelligence (AI) players with sufficient strategic capacity to respond intelligently to the players' action is very challenging (Umarov and Mozgovoy 2012). This last point will be especially relevant if the game wants to demonstrate strategic or even perverse behaviours from the patients, staff, or CFO. Can these virtual actors be programmed to break the rules?

Game design

The game was developed over a 6-month period by a team of public managers, researchers, healthcare managers from the local University Medical Centre, and

programmers from the computer science department. The chief challenge was to translate nuanced research findings into binary computer algorithms, forcing the team to formulate precise assumptions about how performance management systems work. These assumptions cover three elements: (a) modelling the basic operations of a hospital department, (b) identifying the different stakeholder demands and preferences, and (c) specifying the availability and effect of different performance management instruments (see Table 1). As is discussed in more detail below, the assumptions made were a combination of insights from the literature, operational experiences from the University Medical Centre, and decisions to amplify certain performance management effects for educational purposes. Importantly, these assumptions are not posited as paramount truths; the game debrief with the players specifically dedicates time to critically reflect on the extent to which these assumptions are valid in reality and across different public-sector domains or international contexts.

Modelling the activities of a hospital department

The first set of assumptions had to model the operational processes of a hospital, without oversimplifying the complexities of a professional service organization (see Figure 2). Assuming that the actions of professionals are key to the performance of professional

Table 1. Key assumptions in the design of the hospital management game.

| Game component	Key assumptions
(a) Modelling the activities of a hospital department	1. The performance of a hospital department is determined by the time allocation of staff across different activities 2. The different departmental activities are connected to each other, forcing trade-offs between performance parameters 3. Without management intervention, staff will first prioritize patient treatment, then their compulsory administration, then team coordination, and finally personal development time
(b) Identifying different stakeholders and their performance demands	4. A hospital department has to satisfy multiple stakeholders whose satisfaction depends on different parameters 5. Managers can get ongoing feedback from stakeholders about their performance through multiple information sources 6. Without management intervention, most of the stakeholders will not be satisfied with the performance of the department
(c) Operationalizing diverse set of performance instruments and their effects	7. Managers can employ different performance tools, such as targets, rankings, and learning, to influence staff behaviour 8. Targets stimulate staff to achieve the set standard, rankings encourage staff to pursue ever-increasing scores on the selected parameter, learning stimulates gradual movement towards a score deemed desirable by staff themselves 9. When given conflicting performance priorities, staff will prioritize targets over rankings and rankings over learning
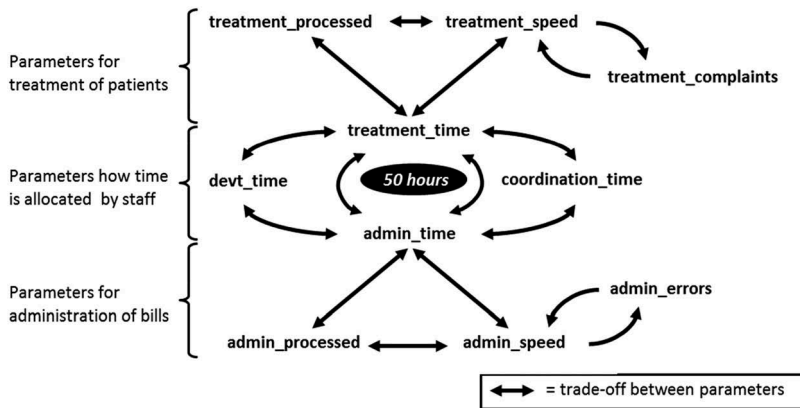
Figure 2. Modelling the key trade-offs in the operation of a hospital department.

Parameters for how time is allocated by staff: The player can influence the allocation of 50 h of staff time available per doctor per week, divided across treatment of patients (treatment_time), administration of billing or financial reporting (admin_time), team coordination between doctors (coordination_time), or personal development time (devt_time). For example, if the player uses the performance instruments to encourage staff to spend 40 h per week on patient treatment, there is only 10 h left for the other tasks.

Parameters for treatment of patients: Extra time for patient treatment could increase the amount of patients processed per hour (treatment_processed), but the output is also determined by the amount of patients seen per hour (treatment_-speed). Furthermore, a higher treatment speed will increase the amount of patient complaints (treatment_complaints). The player can apply performance management instruments to each of the specific parameters, from targeting treatment speed to monitoring patient complaints.

Parameters for administration of bills: Doctors have to process patient bills (or financial records) to ensure the department gets paid for its work. Just like the mechanisms for patient treatment, extra time for administration (admin_time) could lead to more bills being processed (admin_processed), but this is also determined by the amount of bills processed per hours (admin_speed). Furthermore, rushing the billing process will lead to more errors (admin_errors). The player can implement performance instruments to influence staff priorities.

service organizations (Noordegraaf 2015), the focus was on the way medical staff allocated the hours of their work week. Each doctor could allocate 50 hours per week across patient treatment, administration of billing or financial records, team coordination, and personal development time. These different tasks will not exactly reflect the work of all doctors across the world – e.g. publicly versus privately financed healthcare systems will have different billing tasks and incentives for doctors – but we do think that these types of tasks will play some kind of role in most hospitals. The challenge is that there are interactions and trade-offs between the different operational parameters. For example, an increase in time spent on patient treatment (treatment_time) would increase the amount of patients processed per week (treatment_processed), but is also influenced by the amount of patients seen per hour (treatment_speed). Moreover, an increase in treatment speed would lead to more medical errors and patient complaints (treatment_complaints).

At the start of the game, the operational model was purposely loaded with default values creating underperformance. Doctors would spend 25 h per week on patient treatment, 5 h on processing bills, 4 h on team coordination, and 16 h on their personal development time. This state of affairs resulted in the department treating only 1,000 of the 1,500 patients requiring medical attention and billing only 60 per cent of its work. Importantly, the player does not have a full insight in the exact mechanisms of these processes, reflecting the limited overview of managers in complex public organizations. Players have to read management reports and gather stakeholder feedback to see how to improve the running of the department.

Identifying different stakeholders and performance demands

The next design step was to simplify the complex stakeholder field of a hospital department, which includes dozens of actors in reality, while still demonstrating the innate tension between competing stakeholder goals characteristic of the public sector (Speklé and Verbeeten 2014, 131–146). In line with work on the clashing logics governing professional service organizations (Kerpershoek, Groenleer, and de Bruijn 2016, 417–436; Noordegraaf 2015), the medical staff, patients, and chief financial officer are used as representatives of the competing demands faced by hospitals, such as the desire for professional autonomy, good patient outcomes, and cost-effectiveness.

The satisfaction of each stakeholder was tied to different parameters within the model. The medical staff could be satisfied by maximizing time for patients and personal development (high treatment_time, high devt_time), minimizing time spent on administration and avoiding excessive coordination time (low admin_time, moderate coordination_time), while increasing treatment quality (low treatment_complaints). Patients could be satisfied by treating all patients requiring care (high treatment_processed), giving plenty of time to each individual patient (low treatment_speed) and reducing complaints (low treatment_complaints). The chief financial officer could be satisfied by increasing the financial health of the department through maximizing productivity and billing (high treatment_processed and high admin_processed), while avoiding accounting mishaps (low admin_errors).

Throughout the game, players can gauge stakeholder satisfaction by talking to staff around the departmental coffee machine, browsing social media reports from patients, and reading management letters from the CFO. The statements of approval or disproval in these sources are tied to different satisfaction thresholds. The comments from the CFO, for example, would range from 'very satisfied' and 'satisfied' to 'unsatisfied' or 'very unsatisfied' depending on the specific number of bills processed. Exemplifying the contested nature of performance management in the public sector, the game makes it difficult to please all stakeholders. For example, doctors dislike spending time on processing bills, while the CFO would be happy with the extra revenue. It is possible in rare circumstances to get all stakeholders to 'satisfied' (Mikkelsen 2018, 1–21), but that takes the balanced application of different performance management instruments.

Operationalizing different performance instruments and effects

The set of performance management instruments had to represent the multiple forms of performance metrics available, while still giving players a manageable set of options. The game offers players the option to focus on different parameters (time spent on activities, amount of patients processed, complaints, etc.). They then choose to affect the parameters with one of three instruments: *Targets* (setting a threshold value to be achieved), *rankings* (publicly comparing the performance of different doctors), or *intelligence/learning* (generate data for staff discussions without direct consequences (Hood 2012). Finally, players can choose to implement performance metrics through bottom-up involvement of staff in developing the indicators or opt for more top-down management styles. On the whole, the actions of the player are not just about precise performance measurement, but striving for good performance management (Gerrish 2016).

Each of these different instruments has a different impact on the behaviour of staff, especially if they are combined with each other. Targets, for example, lead to aggressive increase in production towards achieving the set target over two turns, rankings lead to ever increasing investment in the selected parameter, while intelligence creates steady and continuous rather than spectacular improvement. For example, setting a target for processing more patients would lead to doctors aggressively re-allocate time to meet the set standard; introducing rankings on this parameter would lead doctors to compete in processing ever more patients; while opting for learning would slowly increase the patients processed.

Combining these tools with different styles generates different types of effects over time. If the player sets clashing priorities, for example, a high target for patients processed in combination with a ranking system for patient complaints, staff will first attempt to meet the target, before committing to rankings or intelligence. Moreover, in those circumstances, staff would prioritize patient treatment over the other tasks such as administration, team coordination, and personal development time. The challenge for the player is to grasp the unpredictable effects of the performance management instruments. To aid the players, they can trace the performance instruments implemented and the results achieved per quarter through an overview board (see Figure 3).

Tracking pathways of players

The game gathers different types of data which can be used to survey the different pathways players pursue to navigate the performance management system. Players are asked to sign a privacy statement, explaining that their data are collected anonymously and will not be shared with third parties or their teachers. A brief survey at the beginning of the game collects information about age, nationality, education, professional background, and attitudes to public service (using the value statements gleaned from the
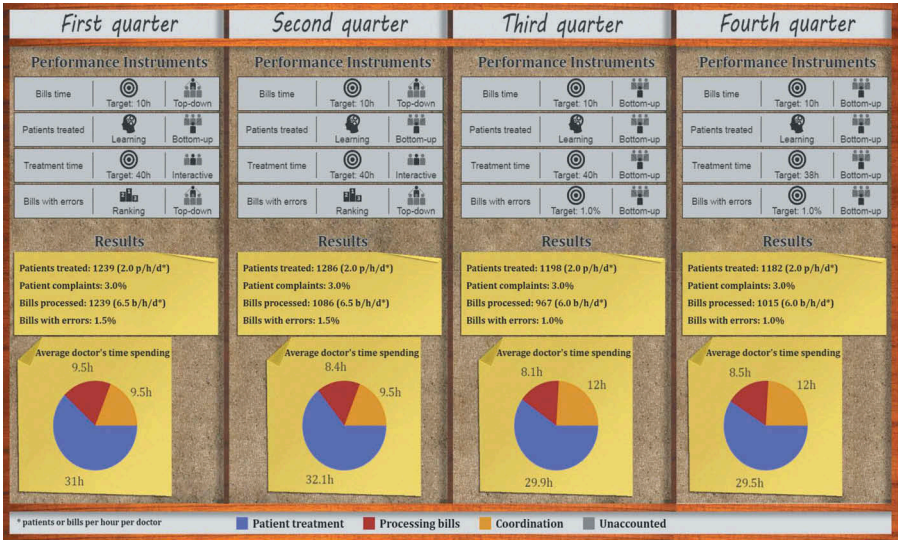

Figure 3. Bulletin board showing performance instruments applied and results achieved per quarter.

European-wide COCOPS public management survey conducted by Van de Walle and Hammerschmid). During the game, data are collected about the performance reports players use, the performance instruments they opt to employ, and the satisfaction level among the different stakeholders. During the preliminary test phases, about sixty different players used the game. In its final form, the game has been played by thirty-three players. As more generations of students will play the game, there will be sufficient data later to conduct full statistical analysis of the different pathways and choices. For now, we can demonstrate the type of insights to be gathered so far, much like a proof-of-concept article introducing a validated survey tool.

Figure 4 shows a pathway pursued by a single player, highlighting the chief information sources consulted in each round (be it the letter from the CFO, a coffee chat with the staff, or social media updates from patients) and the performance domains targeted with performance instruments (be it a focus on the administrative or the care side of the department). The player reported in the pre-game survey that he/she prioritized quality over costs, state over market provision, and achieving results over following rules. The player started quarter 1 with reading the letter from the CFO and introducing targets to reduce the amounts of errors in the bills administration and increase the amount of patients seen per hour. In quarter 2, the player ignored the extra information sources and proceeded to directly increase the targets for both administration and care. (If the player had talked to the doctors, he/she would have learned that they were pushed for time and resolved to prioritize care work over administration tasks.) An angry note from the CFO at the start of quarter 3 led to some extra top-down pressure on the doctors to catch up with their administration. At the end of quarter 4, the patients were relatively happy (as many were seen on time), while the doctor is grumpy (about all the compulsory administration work), and the CFO is angry (about the backlogs in billing).
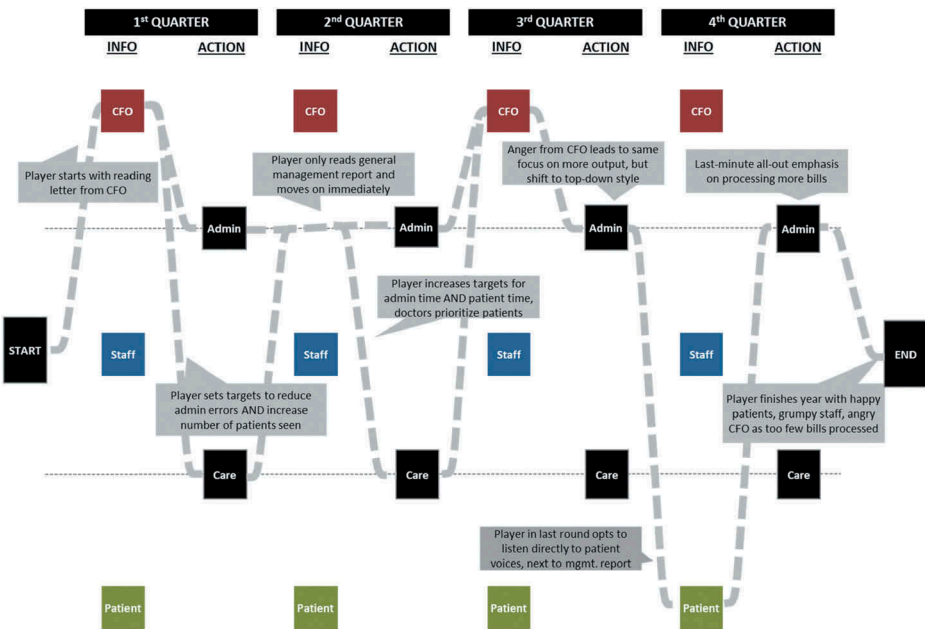

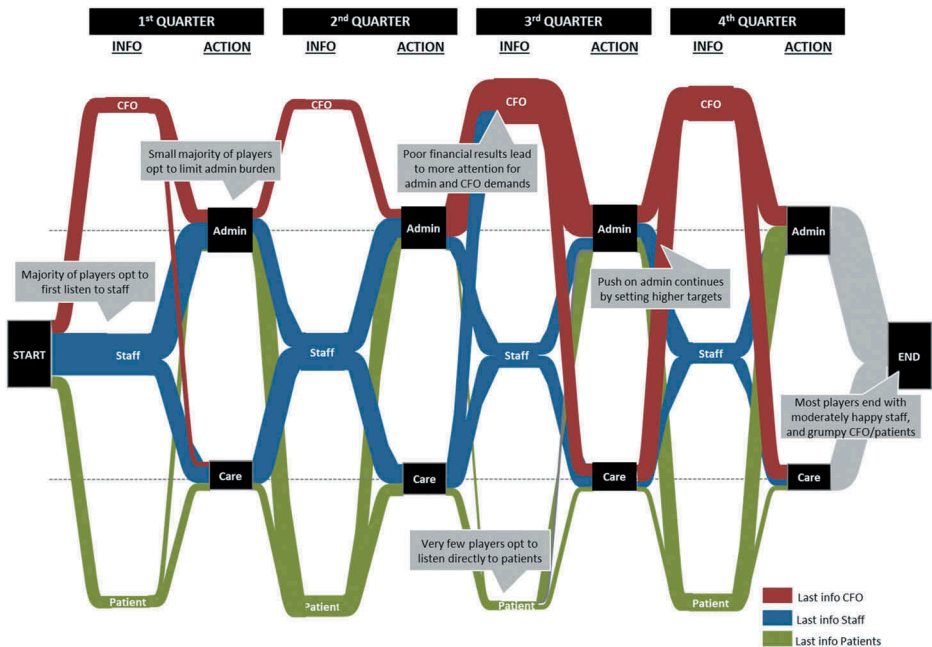Figure 4. Individual pathway through game.

Figure 5. Pathways through performance management system of 33 players.

Figure 5 shows the pathways followed by the thirty-three players which completed the final version of the game. This visualization simplifies some of the choices, but does illustrate how the game data provide an overview of the different pathways. A majority of the actors opted to listen to staff at the start of quarter 1, after which most of them reduced the administrative load on doctors. Most players continued to listen mainly to staff in quarter 2, but some were alerted to the CFO's growing anger about poor financial results, re-introducing administration targets in the later rounds. Most players ended up with moderately happy staff, rather dissatisfied patients, and a decidedly unhappy CFO. There were notable exceptions; some players choose an extreme path by favouring one stakeholder perspective at the expense of others, other players set a target system in quarter 1 and refused to change it, while yet others invested a lot of time each round to read all the information and calibrate the performance instrument accordingly. The current dataset is yet too small for a full analysis of all the different pathways and linking these to player personal backgrounds, but there will be a richer set of observation available when sufficient students and practitioners have played the game over the next two years.

Reflection

Insights for students

The initial tests of the game provide a first indication of its ability to spread insights into the role of diversity, complexity, and contingency in performance management systems. A total of 90 students have played the game so far: A very early version of the game was tested with five PhD students in public management and five medical doctors. The

consecutive test phases were played in different sessions by 25 public management master students, 20 executives taking a master's degree, and 12 mid-level practitioners with mainly financial-management backgrounds. The final version presented here was played by 33 further players from the Netherlands, Germany, Finland, and Spain.

Appreciating diversity

The game design offering multiple information sources and instruments provides players with a first taste of the diversity involved in performance management. Post-game debriefings with players indicated that the forced choice between different information sources and the opportunity to experiment with targets, rankings, and learning techniques helped players to recognize the richness of instruments and techniques on offer. The menu of performance instruments in the game is necessarily highly simplified, but still offered a basis for a wider discussion about the different tools on offer. One criticism was that the individual effects of the different instruments could not be easily gauged by the players, as the underlying interaction effects made it difficult to chart the precise impact of individual interventions. In order to not give a false pretence of simplicity, the decision was made to retain this complexity within the game, but to reveal and discuss the exact mechanisms of the different instruments in the debriefing.

Appreciating complexity

The post-game discussions with players indicated some further insights into the complex effects performance metrics can have on staff behaviour. The game was deliberately programmed to highlight the difficulty of pursuing competing goals. Although the effects may be exaggerated or simplified, the underlying mechanisms are realistic and important for students to appreciate. Again, the learning effect was only achieved by explicitly discussing these assumptions in the post-game briefing and reviewing to what extent they are valid in reality.

One significant limitation in the realism of the game is that the model does capture some unpredictable effects, but does not cover outright perverse effects of performance management such as data distortion, lying, or sabotage (Bevan and Hood 2006). All the reports the players see in the game contain accurate data, while such data would be hard to obtain, hard to measure, or hard to trust in reality. Post-game discussions of test versions revealed that there was already so much information to comprehend and that adding false information would have made the game too difficult. The staff are programmed to pursue their own priority for patient care when faced with conflicting orders, but will not outright defy management instructions. As reported by other authors working in the computer science, it is still too hard to program AI players with sufficient strategic insights to defy game orders or to break the rules (Umarov and Mozgovoy 2012). This game is therefore more about the difficulty of balancing performance metrics, than about the danger of believing performance metrics.

Appreciating contingency

The initial tests also indicate that simulation games helped players to engage with the contingency problems of performance metrics. As players proceeded through the four rounds of the game, they came to appreciate how stakeholders can contradict themselves and how a continuous adjustment of instruments over time may be

necessary. These key assumptions are also revealed and discussed in the debrief, inviting students to reflect on their validity across different organizations or jurisdictions. As the game was programmed to reflect the difficulties in satisfying all stakeholders, it invited particular discussion about the art of balancing preferences in different contexts.

A question was added to the debrief package challenging students and practitioners to think about the applicability of their in-game experiences to different countries and policy domains; would they have experienced the same thing in the management of a police department versus a tax office, a European versus a North American government agency? For practitioners, contrasting the in-game experiences to their experiences in their own real-life organizations was also instructive. To what extent can some of the exaggerated effects in the game also be observed in their organization? What worked in the game that could never work in real life? Above all, this discussion of context must ensure that players do not walk away with a false trust in 'one-best-way' and continue to look for the precise dynamics of different contexts and circumstances.

Insights for researchers

The game design process highlighted various opportunities and challenges for the research of performance management systems. The combination of the pre-game survey and the data on the player choices will provide a novel and rich source of data to track the use of performance information, selection performance instruments, and balancing stakeholder interests. It may generally be interesting to start thinking about the individual response to performance management systems as pathways across parkours, where people have to pragmatically choose one way around a problem. Such a perspective would recognize the complexity and ambiguity of performance management systems, while at the same time acknowledging that managers still have to make practical choices to navigate the system. This perspective and the subsequent data will have to be analysed and discussed in further detail when much larger groups of people have played the game.

The game design process also provided an opportunity to assess the performance management literature from a fresh perspective. The challenge of translating nuanced research findings into the crisp assumptions required for the computer algorithms was both humbling and instructive. This exercise confirmed that we know quite a bit about the different strategies and responses which can be observed in a performance management system, but also showed what the discipline does not know with enough precision. For example, the programmers required a great amount of detail about the decisions of staff: What activities would staff prioritize? How do targets work out over time? We did not have all the academic sources to back up our choices, relying instead on the experience of the medical managers on our team. It will be interesting to examine the answers to these questions in a more scientific manner and investigate how they would play out in different policy domains. We here assumed that staff would prioritize patient care above other activities if given a choice. A management game for another professional service organization, such as a school or law court, may contain similar dynamics, while a management game for a tax agency or ministry may look very different.

The design experience also highlights the limitations of serious games as a research tool. The most important struggle was incorporating the perverse behaviours of people

within performance management systems. The game does include unintended effects, e.g. the push on one performance dimensions adversely affects other outcomes, but does not really go into perverse effects such as output distortion or falsifying data. The play tests showed that the game was already difficult enough for the players without the data on the management report being incorrect, even though that would have been more realistic. Similarly, the game does little to explore the role of relationships between the actors, such as the waxing and waning of trust between management and staff. It proved too difficult for this relatively small game to equip the doctors, patients, and CFO with sufficient AI to craft their own hidden agendas (as predicted by game scholars Umarov and Mozgovoy 2012). If a game has the explicit purpose to investigate distortion and distrust, it might make more sense to create a computer game where humans battle each other.

Conclusion

Computer simulation games provide a valuable additional tool for both teaching and researching the dynamics of performance management systems. The game developed provides some tentative insights into how serious games can be used to display the range of instruments available, trace the often unpredictable effects of management interventions, and review the applicability of instruments across different contexts. However, the inability of the game to capture vital issues such as perverse effects also shows that games cannot replace textbooks or classroom teaching: only the combination of individual game play and offline debrief provides the best learning outcomes. Similarly, games can be used as an additional research method to scrutinize individual choices in navigating performance management systems. In particular, we could gain insights into how individuals use performance information, select performance instruments, and balance the interests of different stakeholders. Again, as serious games struggle to capture some of the social dynamics of performance management systems, they should therefore be paired with other research methodologies.

Disclosure statement

Funding

Notes on contributors

Scott Douglas is Assistant Professor of Public Management at the Utrecht University School of Governance. His research focuses on defining and explaining government success in an age of networks and complexity, frequently collaborating with practitioners to assess challenging domains such as counterterrorism, public health and education.

Christopher Hood is visiting professor at the Blavatnik School of Government and emeritus fellow of All Souls college Oxford. He researches public management, performance management, and budget management.

Tom Overmans is Assistant Professor at the Utrecht School of Governance. He connects the academic fields of 'Public Management' and 'Public Finance'. His knowledge about Public Financial Management stems from his experience as practitioner and scholar.

Floor Scheepers is Professor of Innovation in Mental Health Care and Head of the Psychiatry department of the University Medical Centre Utrecht. She focuses on learning, innovation, and technological breakthroughs which enable better patient-centred care.

References

Angiola, N., and P. Bianchi. 2015. "Public Managers' Skills Development for Effective Performance Management: Empirical Evidence from Italian Local Governments." *Public Management Review* 17 (4): 517–542. doi:10.1080/14719037.2013.798029.

Bevan, G., and C. Hood. 2006. "What's Measured Is What Matters: Targets and Gaming in the English Public Health Care System." *Public Administration* 84 (3): 517–538. doi:10.1111/j.1467-9299.2006.00600.x.

Bouckaert, G., and J. Halligan. 2006. "Performance and Performance Management." In *Handbook of Public Policy*, edited by G. Peters and J. Pierre, 443–459. London: Sage.

Calciolari, S., A. Prenestini, and F. Lega. 2018. "An Organizational Culture for All Seasons? How Cultural Type Dominance and Strength Influence Different Performance Goals." *Public Management Review* 20 (9): 1400–1422.

Calvillo-Gámez, E., J. Gow, and P. Cairns 2011. "Introduction to Special Issue: Video Games as Research Instruments." *Entertainment Computing* 2 (1): 1–2.

Connolly, T. M., E. A. Boyle, E. MacArthur, T. Hainey, and J. M. Boyle. 2012. "A Systematic Literature Review of Empirical Evidence on Computer Games and Serious Games." *Computers & Education* 59 (2): 661–686. doi:10.1016/j.compedu.2012.03.004.

Cooper, S., F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A Leaver-Fay, D Baker, Z. Popović. 2010. "Predicting Protein Structures with a Multiplayer Online Game." *Nature* 466 (7307): 756. doi:10.1038/nature09172.

Crookall, D. 2010. "Serious Games, Debriefing, and Simulation/Gaming as a Discipline." *Simulation & Gaming* 41 (6): 898–920. doi:10.1177/1046878110390784.

De Smale, S., T. Overmans, J. Jeuring, and L. van de Grint 2015. "The Effect of Simulations and Games on Learning Objectives in Tertiary Education: A Systematic Review." In *International Conference on Games and Learning Alliance*, 506–516. Cham: Springer.

Emerson, K., and T. Nabatchi. 2015. *Collaborative Governance Regimes*. Georgetown: Georgetown University Press.

Gerrish, E. 2016. "The Impact of Performance Management on Performance in Public Organizations: A Meta-Analysis." *Public Administration Review* 76 (1): 48–66. doi:10.1111/puar.12433.

Hood, C. 2012. "Public Management by Numbers as a Performance-Enhancing Drug: Two Hypotheses." *Public Administration Review* 72 (s1): S85-S92. doi:10.1111/j.1540-6210.2012.02634.x.

Kerpershoek, E., M. Groenleer, and H. de Bruijn. 2016. "Unintended Responses to Performance Management in Dutch Hospital Care: Bringing Together the Managerial and Professional Perspectives." *Public Management Review* 18 (3): 417–436. doi:10.1080/14719037.2014.985248.

Kusurkar, R. A., G. Croiset, and O. T. J. Ten Cate. 2011. "Twelve Tips to Stimulate Intrinsic Motivation in Students through Autonomy-Supportive Classroom Teaching Derived from Self-Determination Theory." *Medical Teacher* 33 (12): 978–982. doi:10.3109/0142159X.2011.599896.

Laird, J. E. 2002. "Research in Human-Level AI Using Computer Games." *Communications of the ACM* 45 (1): 32–35. doi:10.1145/502269.502290.

Lewandowski, M. 2018. "Public Managers' Perception of Performance Information: The Evidence from Polish Local Governments." *Public Management Review*. doi:10.1080/14719037.2018.1538425.

Mayer, I., G. Bekebrede, C. Harteveld, H. Warmelink, Q. Zhou, T. van Ruijven, I. Wenzler, R. Kortmann, and I. Wenzler. 2014. "The Research and Evaluation of Serious Games: Toward a Comprehensive Methodology." *British Journal of Educational Technology* 45 (3): 502–527. doi:10.1111/bjet.2014.45.issue-3.

Meier, K., S. C. Andersen, L. J. O'Toole Jr, N. Favero, and S. C. Winter. 2015. "Taking Managerial Context Seriously: Public Management and Performance in US and Denmark Schools." *International Public Management Journal* 18 (1): 130–150. doi:10.1080/10967494.2014.972485.

Mikkelsen, M. F. 2018. "Do Managers Face a Performance Trade-Off? Correlations between Production and Process Performance." *International Public Management Journal*: 21 (1): 53–73.

Moore, M. 2013. *Recognizing Public Value*. Cambridge: Harvard University Press.

Moynihan, D. P., and A. Kroll. 2016. "Performance Management Routines that Work? an Early Assessment of the GPRA Modernization Act." *Public Administration Review* 76 (2): 314–323. doi:10.1111/puar.12434.

Narayanasamy, W., K. K. Wong, C. C. Fung, and S. Rai. 2006. "Distinguishing Games and Simulation Games from Simulators." *Computer Entertainment* 4 (2): 9. doi:10.1145/1129006.

Nõmm, K., and T. Randma-Liiv. 2012. "Performance Measurement and Performance Information in New Democracies: A Study of the Estonian Central Government." *Public Management Review* 14 (7): 859–879. doi:10.1080/14719037.2012.657835.

Noordegraaf, M. 2015. *Public Management: Performance, Professionalism and Politics*. London: Palgrave Macmillan.

Overmans, T., W. Bakker, Y. van Zeeland, G. van der Ree, J. Jeuring, M. van Mil, and W. Dictus 2017. *The Value of Simulations and Games for Tertiary Education*. https://dspace.library.uu.nl/bitstream/handle/1874/327311/2015_017.pdf

Palmunen, L.-M., E. Pelto, A. Paalumäki, and T. Lainema. 2013. "Formation of Novice Business Students' Mental Models through Simulation Gaming." *Simulation & Gaming* 44 (6): 846–868. doi:10.1177/1046878113513532.

Pollitt, C. 2018. "Performance Management 40 Years On: A Review. Some Key Decisions and Consequences." *Public Money & Management* 38 (3): 167–174. doi:10.1080/09540962.2017.1407129.

Skelcher, C., and S. R. Smith. 2015. "Theorizing Hybridity: Institutional Logics, Complex Organizations, and Actor Identities: The Case of Nonprofits." *Public Administration* 93 (2): 433–448. doi:10.1111/padm.12105.

Smith, S. P., and D. Trenholme. 2009. "Rapid Prototyping a Virtual Fire Drill Environment Using Computer Game Technology." *Fire Safety Journal* 44 (4): 559–569. doi:10.1016/j.firesaf.2008.11.004.

Speklé, R. F., and F. H. Verbeeten. 2014. "The Use of Performance Measurement Systems in the Public Sector: Effects on Performance." *Management Accounting Research* 25 (2): 131–146. doi:10.1016/j.mar.2013.07.004.

Talbot, C. 2008. "Performance Regimes—The Institutional Context of Performance Policies." *International Journal of Public Administration* 31 (14): 1569–1591. doi:10.1080/0190069080219943.

Ten Cate, O., L. Snell, K. Mann, and J. Vermunt. 2004. "Orienting Teaching toward the Learning Process." *Academic Medicine: Journal of the Association of American Medical Colleges* 79 (3): 219–228. doi:10.1097/00001888-200403000-00005.

Torenvlied, R., A. Akkerman, K. J. Meier, and L. J. O'Toole. 2013. "The Multiple Dimensions of Managerial Networking." *The American Review of Public Administration* 43 (3): 251–272. doi:10.1177/0275074012440497.

Umarov, I., and M. Mozgovoy. 2012. "Believable and Effective AI Agents in Virtual Worlds: Current State and Future Perspectives." *International Journal of Gaming and Computer-Mediated Simulations* 4 (2): 37–59. doi:10.4018/IJGCMS.

Van Dooren, W., G. Bouckaert, and J. Halligan. 2015. *Performance Management in the Public Sector*. 2nd ed. London: Routledge.

Vasiliou, C., and P. Cairns 2011. *Evaluating Human Error through Video Games*.

Verhoest, K., and J. Wynen. 2016. "Why Do Autonomous Public Agencies Use Performance Management Techniques? Revisiting the Role of Basic Organizational Characteristics." *International Public Management Journal* 21 (4): 619–649.