

Output measurement in professional public organizations: insights from European justice systems

European
justice systems

1

Petra Pekkanen and Timo Pirttilä
LUT University, Lappeenranta, Finland

Received 29 June 2021
Revised 9 February 2022
19 August 2022
Accepted 27 August 2022

Abstract

Purpose – The aim of this study is to empirically explore and analyze the concrete tasks of output measurement and the inherent challenges related to these tasks in a traditional and autonomous professional public work setting – the judicial system.

Design/methodology/approach – The analysis of the tasks is based on a categorization of general performance measurement motives (control-motivate-learn) and main stakeholder levels (society-organization-professionals). The analysis is exploratory and conducted as an empirical content analysis on materials and reports produced in two performance improvement projects conducted in European justice organizations.

Findings – The identified main tasks in the different categories are related to managing resources, controlling performance deviations, and encouraging improvement and development of performance. Based on the results, key improvement areas connected to output measurement in professional public organizations are connected to the improvement of objectivity and fairness in budgeting and work allocation practices, improvement of output measures' versatility and informativeness to highlight motivational and learning purposes, improvement of professional self-management in setting output targets and producing outputs, as well as improvement of organizational learning from the output measurement.

Practical implications – The paper presents empirically founded practical examples of challenges and improvement opportunities related to the tasks of output measurement in professional public organization.

Originality/value – This paper fulfils an identified need to study how general performance management motives realize as concrete tasks of output measurement in justice organizations.

Keywords Performance measurement, Public sector organizations, Professions, Justice system, Output measurement

Paper type Research paper

1. Introduction

Performance management (PM) systems are increasingly implemented in different sectors of professional public sector organizations (e.g. Barbato and Turri, 2017; de Bruijn, 2011; Goh *et al.*, 2015; Vogel, 2013). Despite the widespread and worldwide use of the systems, there still exists several barriers, challenges and problems in the effective utilization of performance measurement systems in these organizations (Barbato and Turri, 2017; Jääskeläinen and Roitto, 2014; Goh *et al.*, 2015). The identified and most studied distinct characteristics and challenges are connected to the output versus outcome measurement, resource allocation and budgeting practices, the number of stakeholders involved and to the unintended behavior and effects caused by the implemented PM system (e.g. Barbato and Turri, 2017; de Bruijn, 2011; Johansson, 2015; Radnor, 2008; Rajala *et al.*, 2018; Goh *et al.*, 2015; Van Thiel and Leeuw, 2002).

A distinct research area is the relationship between output and outcome measurement in professional public sector organizations (Rajala *et al.*, 2018). Output measurement should not



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

International Journal of
Productivity and Performance
Management
Vol. 73 No. 11, 2024
pp. 1-19
Emerald Publishing Limited
1741-0401
DOI 10.1108/IJPPM-05-2021-0279

be forgotten in the improvement of performance measurement systems. As outcomes can be seen to cover the long-term effects of performance management (Førsund, 2012; Mandl *et al.*, 2008; Rajala *et al.*, 2018), it is important to make sure that output measures and measuring practices are appropriate and comprehensively incorporate the important components of the organization's performance. This requires that output and outcome measurement are not understood as opposites, but rather mutually complementary and important parts of the overall PM system (Linna *et al.*, 2010). In professional service organizations, the outputs are also usually intangible and more difficult to quantify than in a manufacturing environment. This means that the different challenges related to performance measurement can only be solved after the outputs and their effects on outcomes have been properly defined (Jääskeläinen and Lönnqvist, 2011). Hence, it is important that the improvement needs and opportunities of output measurement are not forgotten, but instead studied and analyzed further in different organizational contexts in order to improve the transformation of outputs into the desired outcomes and avoid possible negative effects on operations.

The various dimensions of quality are usually well established in the processes, management practices and professional ethics of the organizations (Pekkanen, 2011), whereas efficiency issues are considered irrelevant and contradictory to the overall purpose of the organization (de Bruijn, 2011; Johansson, 2015). However, public organizations are dealing with problems connected to efficiency and productivity (de Bruijn, 2011; Dobija *et al.*, 2019; Elg *et al.*, 2013), and public managers are facing increasing pressure to hold down public expenditure while improving service quality (Björk *et al.*, 2014). The increasing pressures to improve efficiency and accountability also highlights the importance and the need to analyze and improve the output measurement systems used in the organizations (Jääskeläinen and Lönnqvist, 2011; Mandl *et al.*, 2008). There exists a clear need to study the possibilities to design suitable and more comprehensive output indicators and measurement systems without compromising quality, equality and satisfaction of citizens or other professional work standards (see, e.g. Jääskeläinen and Lönnqvist, 2011). This calls for empirical analysis on the motives, tasks, and main improvement needs of the implemented output measures.

Implementing performance management systems often causes tension between managers and professionals, making the autonomous professionals to reject or abuse the proposed indicators (de Bruijn, 2011). Professionals opposing or gaming the indicators is usually a consequence of choosing indicators based on practicality, leading to the oversimplification of the complex work of professionals (de Bruijn, 2011; Johansson, 2015). It has been argued that PM in professional public sector organizations is viewed too narrowly, focusing on control and accountability and not enough on the motivation, improvement, and learning tasks of measurement systems (Radnor, 2008). It should be remembered that measurement information has several important aspects to different stakeholders (e.g. society, management, professionals), and include these aspects into the systems. By studying and analyzing output measurement from the perspective of different motives and purposes, it would be possible to better incorporate incentives and feedback for improvement, celebrate and promote learning from the measurement information, facilitate self-management of professionals, and reduce the perverse effects and negative attitudes related to PM systems (Behn, 2003; Johansson, 2015; Speklé and Verbeeten, 2014). Research on measurement motives has produced lists and characteristics of the different purposes of PM systems (e.g. Behn, 2003; Speklé and Verbeeten, 2014). However, there is a lack of empirical studies on how these different motives manifest in concrete output measures implemented in professional public organizations and how the used output measures support the achievement of the different motives from the perspective of different stakeholders.

The aim of this study is to empirically explore and analyze the main tasks of output measurement and the inherent challenges related to the tasks in a very traditional and autonomous professional public work setting – the judicial system. In justice organizations,

the challenges surrounding performance measurement is highlighted because the emphasis on efficiency and accountability is considered a threat to the autonomy and impartiality of the judges (De Santis and Emery, 2017; Lienhard *et al.*, 2012). The outcomes of courts are especially difficult to define precise (Contini *et al.*, 2014; Vecchi, 2018), making the comprehensiveness and impressiveness of the output measures especially important.

The research questions of the study are:

- (1) How are the general performance management motives realized as concrete tasks of output measurement in justice organizations?
- (2) What are the main challenges in carrying out these different tasks of output measurement?

The analysis is based on a categorization of general performance measurement motives (control-motivate-learn) and main stakeholder levels (society-organization-professionals). The analysis is exploratory and based on secondary source of evidence: materials and reports produced in two performance improvement projects in European judicial systems. The analyzed material is based on interviews and expert workshops conducted in 14 European countries and their judicial organizations.

Section 2 presents a literature review concerning the distinct characteristics and motives of performance management in professional public sector organizations. Section 3 covers the methodology of the study. Section 4 presents the main findings of the study. Finally, in section 5, discussion on the results and concluding remarks are presented.

2. Literature review

The public sector in many countries has undergone a range of New Public Management (NPM) reforms over the last decades to become more “business-like” with a greater emphasis on results and accountability (see, e.g. Charbonneau *et al.*, 2015; McGeough, 2015; Pollitt and Bouckaert, 2000). Hood (1991) and Van Dooren *et al.* (2015) conclude in their studies that performance measurement and control, especially output measurement, are central doctrinal components in NPM. Many professional public organizations have introduced performance measurement systems in order to meet the increasing requirements concerning efficiency, transparency and performance accountability (e.g. Barbato and Turri, 2017; de Bruijn, 2002; Dobija *et al.*, 2019; Elg *et al.*, 2013; Linna *et al.*, 2010; Radnor and McGuire, 2004; Tabi, 2013; Vogel, 2013). However, it is argued that, the literature on the performance measurement systems in professional public sector organizations needs more detailed empirical analysis (see, e.g. Jääskeläinen and Lönnqvist, 2011).

The basic reason for using performance measurement is to improve goal achievement by shaping behavior within the organization and to act as the basis for internal and external accountability (Johansson, 2015). However, performance measures should also be incentives to promote improvement, learning and motivation in the organization (de Bruijn, 2002; Radnor, 2008). In addition, it should be noted in terms of PM systems, that public sector performance is multidimensional, including explicit considerations of quality and quantity of outputs, service outcomes, and citizen and user satisfaction (Charbonneau *et al.*, 2015).

2.1 Performance measurement in professional public organizations

The most studied distinctive characteristics of performance measurement in professional public organizations are connected to (1) the output vs. outcome measures and measurement, (2) the issues surrounding resource allocation and budgeting practices, (3) the challenges connected to the number of stakeholders involved in measurement, and (4) the unintended effects of the implemented systems (e.g. de Bruijn, 2011; Johansson, 2015; Radnor, 2008; Rajala *et al.*, 2018; Goh *et al.*, 2015; Van Thiel and Leeuw, 2002).

Outputs describe what the organization does in using resources to produce direct outputs, and outcomes describe what the direct and indirect effects of the outputs are (e.g. Charbonneau *et al.*, 2015; Rajala *et al.*, 2018). A typical feature in the relationship between outputs and outcomes is that the transformation from outputs to outcomes includes various external factors. Therefore, the different transmission effects are hard to distinguish and isolate, and the process is not fully controllable by the organization. It is also challenging to disentangle the effects of different outputs on the outcome. There are usually delays between the implementation of output measures and their impact on the outcome (Førsund, 2012; Mandl *et al.*, 2008).

It is still extremely complicated to establish an outcome-oriented measurement system for public sector needs, purposes, and practices (Rajala *et al.*, 2018). Some outcomes cannot be measured directly, and some cannot be measured at all. Outcomes usually reflect values like quality and satisfaction, which are hard to define and measure (Rajala *et al.*, 2018). As more information and more time are used to collect outcome data, it may be even harder to point out the factors causing the particular outcome (Lowe, 2013). However, performance management is unlikely to be effective if it does not include both outputs and outcomes as part of an integrated performance information and evaluation framework (McPhee, 2005). As outcomes cover the long-term effects of performance management (Førsund, 2012; Mandl *et al.*, 2008; Rajala *et al.*, 2018), output measures and measuring practices need to be appropriate and comprehensively incorporate the important components of the organization's performance. This requires that the output and outcome measurement are not understood as opposites, but rather mutually complementary and important parts of the overall performance management system (Linna *et al.*, 2010). Improving output measurement should not be forgotten in the improvement efforts of professional public organizations.

Usually in the public sector, performance measures are highly related to budget and resource allocation decisions. This makes the design of the systems to also include a political dimension with large impact from the society driven by the need for equality (de Bruijn, 2002; Linna *et al.*, 2010; Goh *et al.*, 2015). The use of measures in resource allocation and strategic planning have been widely studied (e.g. Jämskeläinen and Roitto, 2014). The most common approach to allocating resources in the public sector is performance-based funding (PBF). Various alternatives of this system have been developed in different countries (see, e.g. Hur, 2018). PBF uses specific formulas to tie funding to organizations' performance, which is, again, based on different indicators and targets (Francesconi and Guarini, 2018). In the Western world, a so-called model of management by objectives and results (MBOR) is commonly utilized (see, e.g. Kristiansen, 2017). In MBOR, agencies and individual organizations are given autonomy and flexibility in the use of resources and in choosing means and measures, but they must accept performance contracts, targets, reporting and assessment systems established by the ministries or other governmental institutes (Kristiansen, 2017).

de Bruijn (2002) states several common challenge areas in the financial incentive and pay-for-performance aspects of resource allocation and budgeting practices. There is a possibility that the practices can even lead to punishment of performance. The challenges are connected to shared resource pools, lack of performance transparency and lack of possibilities to reward good performance. When a fixed budget is divided among several organizations and all organizations perform better, it leads to financial sanctions with lower "price-per-product". A transparent and well-performing organization may also be in a vulnerable situation, where investment in increasing efficiency may lower the budget for the next year (whereas an organization not increasing efficiency and offering transparency may be rewarded with equal targets and resources). Rewarding well-performing organizations with resources is not straightforward. Usually additional resources need to be used to help non-performing organizations to guarantee equal service to citizens, sometimes at the expense of the well-performing organizations.

In public organizations, there exists multiple stakeholders with conflicting needs concerning performance. This may be realized as a large number of measures and unfocused purpose of measurement (Boland and Fowler, 2000; Jääskeläinen *et al.*, 2014; Jääskeläinen and Roitto, 2014; Radnor, 2008; Rantanen *et al.*, 2007). Hence, conceptualizations of the performance of an organization may vary depending on the perspective from which it is viewed. Some perspectives are contradictory, creating an especially challenging task for managers as they try to achieve improved results (Charbonneau *et al.*, 2015). Important stakeholder groups include the citizen- and the formal state society, the management of the organizations and the individual professionals. The role of the society is to control that individual organizations produce the outputs and outcomes needed. As the society has a central role, Radnor and McGuire (2004) argue in their study that the role of managers is often more about being administrators than managers, particularly in relation to performance measurement. Managers need to balance between sometimes conflicting goals, required to follow national and local policies while being attentive to the needs of staff and customers during day-to-day service production (Björk *et al.*, 2014).

In professional organizations, the performance measurement practices can lead to different unwanted effects and behavior. These perverse effects are widely referred to in literature (e.g. de Bruijn, 2002; de Bruijn, 2011; Garlatti *et al.*, 2018; Kerpershoek *et al.*, 2016; Radnor, 2008). Typical effects include, for example, different types of strategic behavior (e.g. optimizing the output, concentrating on easier outputs, focusing on performance on unit-level not on organizational level, focusing on short-term targets, misreporting or distorting data, and deliberate under-achieving). Van Thiel and Leeuw (2002) divide the perverse effects to unintended and deliberate. Unintended perverse effects can be caused by insufficient knowledge about performance measurement, and deliberate perverse effects happen through conscious decisions. The characteristics of professional public organizations foster the emergence of perverse effects (Van Thiel and Leeuw, 2002), and this should be considered in the design of the system. It has been noted that commitment towards measures is more widely established when management and other personnel feel they can affect the measurement results (Jääskeläinen and Roitto, 2014).

2.2 Motives of performance measurement in professional public organizations

According to various studies, performance measures should be, among other things, diverse and complementary, objective, informative, causally related, supportive of decision-making and incentive for improvement (e.g. Ittner and Larcker, 2002; Lipe and Salterio, 2000; Malina and Selto, 2004). In general, performance measures should enable transparency, learning, appraising, sanctioning, and comparing (de Bruijn, 2002). Performance measurement systems are also meant to create incentives to align individual goals with the objectives of the organization, provide feedback information concerning the progress towards these objectives, and form the basis for internal and external accountability (Speklé and Verbeeten, 2014).

Public performance management literature also recognizes various uses for performance information and discusses the factors behind the use of performance information (e.g. Behn, 2003; Laihonon and Mäntylä, 2017). Measures act as evidence of effectiveness in improving public accountability and policy decision-making, as well as enable goal setting, resource allocation and budget formulation (Behn, 2003).

Several different categorizations of motives and uses of performance measures can be identified in the literature. For example, Speklé and Verbeeten (2014) introduce three different classifications of performance measurement use: operational use, incentive use and exploratory use. Vecchi (2018) divides the usage into four categories: instrumental use, process use, conceptual/enlightenment use, and symbolic/legitimizing use. Behn (2003) defines and explains eight purposes:

- (1) Evaluate – How well is an agency performing?
- (2) Control – Ensuring that subordinates are doing the right things.
- (3) Budget – On what should the agency spend the public's money?
- (4) Motivate – How to motivate line staff, middle managers, collaborators, stakeholders, and citizens to do things to improve performance?
- (5) Promote – How to convince stakeholders that the agency is doing a good job?
- (6) Celebrate – What accomplishments are worthy of celebrating success?
- (7) Learn – What is working, what is not?
- (8) Improve – What exactly should be done differently to improve performance?

It can be concluded that, in addition to decision support and control, performance information should also have other purposes ([Laihonen and Mäntylä, 2017](#); [Pollitt, 2006](#)). It has been argued that, in general, controlling plays too large of a role at the expense of the other measurement purposes ([Radnor, 2008](#)). To produce different types of positive effects (for example innovations and organizational learning), the development of structures and procedures to enable these effects would be required (e.g. [Kalgin et al., 2018](#); [Laihonen and Mäntylä, 2017](#)).

3. Methods and data

This study is conducted as an empirical content analysis, utilizing data collected from continental European justice organizations. The study is based on two separate research and improvement projects carried out in co-operation between several European research and governmental institutes and co-funded by the European Commission. The empirical content analysis is based on secondary source of evidence, utilizing the written materials and reports produced during the research and improvement projects.

Justice systems can be categorized under broader labels or so called major legal systems: the common law system and the civil law system (sometimes referred to as Romano-Germanic legal system). Continental Europe belongs to the group of civil law system. Central to the civil law system is that laws are based on codification and the judge's role is to apply the law instead of creating it. The civil law tradition also emphasizes the judicial independence of a judge, which is needed to secure the fairness of judgments (see, e.g. [Merryman and Pérez-Perdomo, 2007](#)).

3.1 Data gathering

The analysis is based on a set of written materials and reports produced in two performance improvement projects at the European judicial systems. The authors have been involved in the data gathering and reporting of both projects. The projects investigated issues related to the improvement needs and opportunities of performance and quality assurance practices in different European countries.

The first project “CFM-net – Towards European Caseflow Management development network – Identifying, developing and sharing best practices” was carried out in 2014–2016. The overall objective of the project was to start creating procedures for European co-operation in developing and sharing good practices for process and operations management – namely the “flow” of judicial cases and performance of the organizations. The research team included members from 5 countries: Estonia, Finland, Italy, the Netherlands, and Switzerland. Two reports from the project are included in the data

analysis: “Caseflow Management Handbook – Guide for Enhanced Court Administration in Civil Proceedings” and “Inventory of caseflow management practices in European civil proceedings”. The first report provides a facilitation guide and general analysis, guidelines, and advice for carrying out improvement work in courts. The second report is a collection of tools, practices and solutions applied in different European countries related to process performance improvement. The reports are based on an extensive set of data gathered during the project. The data gathering in the project included interviews and on-site visits in 12 different European countries, four expert workshops, as well as literature reviews and materials related to other court improvement projects. Information about the project and the project reports and materials can be found at: www.lut.fi/web/en/european-caseflow-management-development-network.

The second project “*Handle with Care: assessing and designing methods for evaluation and development of the quality of justice*” was carried out in 2017. The overall objective of the project was to study, analyze and improve the quality of justice in European countries. The project looked at the justice system as a whole, and the procedures aimed at improving the ways in which the system performs, operates, and generates public value. The quality concept in the project was broadly defined as covering practices and procedures of performance management and evaluation. The reports cover “classical” performance management practices, as well as more innovative practices emerging in the countries. The research team included members from 5 countries: Finland, France, Hungary, Italy, and the Netherlands. Four reports from the project are included in the data analysis: “Handle with Care: Assessing and designing methods for evaluation and development of the quality of justice” (summative report of project results), “Comparing the evaluation and development of the quality of Justice in Finland, France, Hungary, Italy and the Netherlands” (compares data collected at national level to identify common trends and diverging paths), “Something good? In search of new practices to improve the quality of justice in EU” (identifies innovative practices and analyzes conditions for a successful implementation at EU level), and “Performance management of courts and judges - organizational and professional learning vs political accountability” (develops a methodological framework for performance and quality evaluation and improvement in courts). The reports are based on an extensive set of data collected during the project. The data gathering in the project included interviews in five countries, two expert workshops, as well as literature reviews and documentary analysis. Information about the project and the project reports and materials can be found at: <https://www.lut.fi/web/en/school-of-engineering-science/research/projects/handle-with-care>.

Table 1 summarizes the materials and reports produced in the projects and their data gathering methods.

3.2 Data analysis

First, an analysis framework was formed based on literature. The analysis framework enabled the categorization of the tasks according to two dimensions: the main motive of the measurement and the main stakeholder level. The aim of this categorization was to create an overall picture of the output measurement and enable detailed analysis on how the general performance measurement motives are realized as concrete tasks from the perspective of different stakeholders.

Three stakeholder groups were included in the data analysis framework: individual (judge), organizational (court) and societal (formal national and governmental justice institutes). The measurement practices were further categorized according to these groups into two measurement levels: measurement tasks at the level of *society – organization* and on the level of *organization-individual*. Measurement tasks set on the level of society-individual

Table 1.
Summary of the
analyzed reports and
data gathering
methods

Project	Reports and materials	Data gathering methods
CFM-net - Towards European Caseflow Management development network – Identifying, developing, and sharing best practices - Duration: 2 years (2014–2016) - Research team: 15 members from 5 countries - Team included: Research institutions, Universities, Ministries	(1) Caseflow Management Handbook – Guide for Enhanced Court Administration in Civil Proceedings (38 pages) (2) Inventory of caseflow management practices in European civil proceedings (57 pages)	(1) Interviews and on-site visits 60 individual interviews (judges, court administrators/ managers) 12 European countries: Austria, Belgium, Czech Republic, Finland, Germany, Estonia, Italy, the Netherlands, Portugal, Slovenia, Spain, and Sweden 7 on-site visits in courts (2) Expert workshops 4 two-day analysis workshops 12 participants (judges, court managers, academics) (3) Literature review
HWC - Handle with Care: assessing and designing methods for evaluation and development of the quality of justice - Duration: 1 year (2017) - Research team: 22 members from 5 countries - Team included: Research institutions and Universities	(1) Handle with Care Assessing and designing methods for evaluation and development of the quality of justice (405 pages) (2) Comparing the evaluation and development of the quality of Justice in Finland France, Hungary, Italy, and the Netherlands (25 pages) (3) Something good? In search of new practices to improve the quality of justice in EU (33 pages) (4) Performance management of courts and judges-organizational and professional learning vs political accountability (33 pages)	(1) Interviews 15 individual interviews (judges, court administrators/ managers, policy makers) 5 European countries: Finland, France, Hungary, Italy, the Netherlands (2) Expert workshops 2 one-day analysis workshops 25 participants (judges, policy makers, court managers, academics) (3) Literature review and documentary analysis

(e.g. measures related to occupational development, position appointment and promotion) were also identified. These were excluded from the categorization as the measures were not related to output, but rather on quality-related issues in professional work of judges.

Three basic measuring motive categories were formed based on the literature review and included in the data analysis framework: control, motivate and learn. The aim of this concise categorization was to form the focal and most central motive categories for the purposes of coherent data analysis. The control category was defined to include also the different aspects of financial performance (e.g. budgeting) and the evaluation of performance. The learn category included also the promote, celebrate and improve motives of measurement. Other categorizations have been utilized in different studies in the field (see, e.g. Behn, 2003; Speklé and Verbeeten, 2014; Vecchi, 2018). These categorizations have been used as a basis in creating adequately simple and distinctive categories for analysis purposes.

The basic measuring roles in the data analysis are defined as the “Principal” (responsible for diagnostic and measurement decisions) and the “Actors” (responsible for the actual operations and performance).

The categories included in the analysis were defines as:

(1) Control

- Purpose: to plan operations and to monitor process and performance
- The principal identifies deflections in performance levels and executes corrective actions (single-loop learning)

(2) Motivate

- Purpose: to encourage and influence actors in improving performance
- The actor decides and implements improvement initiatives and the principal rewards success

(3) Learn

- Purpose: to understand performance improvement opportunities and the factors affecting performance
- All those concerned analyze actual performance and process-related factors. Aiming to find a “common path” for success and promoting a wider application of that path (double-loop learning).

Based on the categorizations along these dimensions, the framework for data analysis included six output measurement categories: control measures for society-organization level and for organization-individual level, motivate measures for society-organization level and for organization-individual level, as well as learn measures for society-organization level and for organization-individual level (see [Figure 1](#)).

In the data analysis process, the output measurement tasks, and the inherent challenges described in the reports were identified, coded, and categorized according to the data analysis framework and analyzed further. Firstly, all tasks related to output measurement were identified, marked in the material, listed and given an initial name. After the tasks were listed

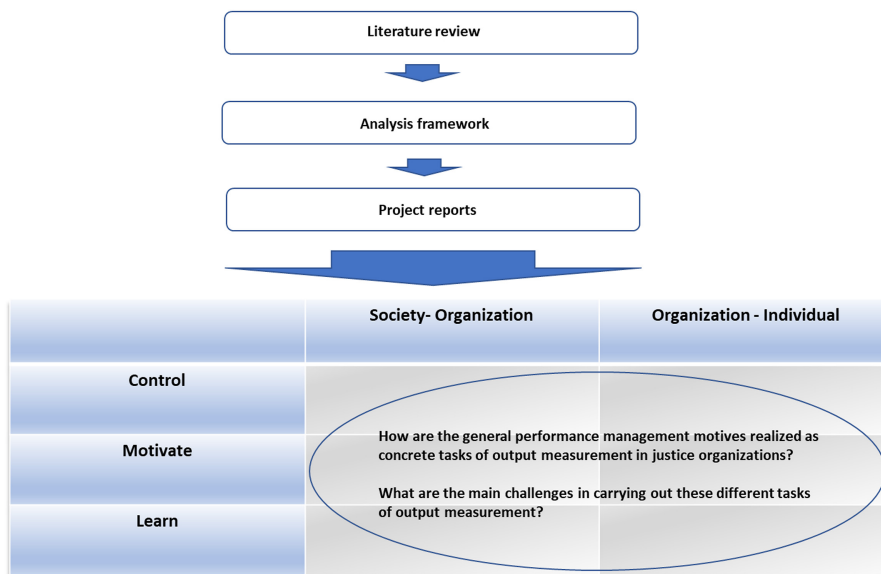


Figure 1.
Summary of the phases
and results of the data
analysis process

and named, their underlying motives and use were analyzed for the categorization purposes. The materials provided detailed descriptions of the underlying aims and the use of the output measures in studied countries and organizations. These descriptions were used as a basis to relate each identified task to the category based on its main motive and stakeholder level. Two researchers first conducted this analysis of the reports and materials separately, after which the analyses and categorizations result were verified, discussed, and combined. At this point the reports were revisited multiple time in order to make sure that all relevant information related to individual tasks were included to the categorization.

In the second phase, the challenges and features of each identified and categorized task were analyzed further. The analyzed material included lot of information from different countries about the features and challenges connected to the output measurement tasks. These were utilized as a source in conducting for each task a summative description of main challenges connected to it. The analyzed tasks and the descriptions were broadened with documented and described examples and experiences. The data analysis was conducted at European level, not separating measurement practices according to different countries.

A summary of the main phases and results of the data analysis process is described in [Figure 1](#).

4. Findings

The main findings related to the tasks in the motive categories and at the two stakeholder levels are summarized in [Table 2](#). The tasks, and the challenges related to them, are described in more detail in the following sub-sections.

4.1 Tasks of output measurement at the society – organization level

The control category at the society - organization measurement level was found to include tasks related to resource management and performance deviation management. Firstly, the control motive is connected to the task of overseeing and balancing workload and resources between individual court organizations. The basic principle connected to the task is the inherent need for objectivity in setting output goals and allocating resources between the organizations. The number of produced outputs is seen as a straightforward and easy measure to base resource allocation decisions on. The goal for the number of outputs and the resources needed are decided in the annual output-based budget and funding negotiations. The challenge inherent to the task is that the court organizations have large variations in their circumstances, operating environments, case-structures and thus in their workloads and resource needs. The differences in the case structure mean that the judicial cases vary in complexity and thus in the amount of time and resources needed for them. The case structure also varies from one year to another and cannot be completely predicted and planned. Therefore, mechanisms are required to proactively detect changes in the workload leading to over- or under-resourcing and the need for re-allocation of resources.

The challenges related to the commensuration of output goals and resources were found to be common in all studied countries. To tackle the challenges related to variations in case-structures and resource needs, different types of weighted caseload systems have been introduced. The systems aim to provide data that are more accurate for goal setting and resource allocation and to provide opportunities to compare the resource utilization of different courts more reliably and in more detail. Even though the need for these types of resource allocation tools is common for all the studied countries, the readiness and level of detail used differ among the countries. In Italy, the need for workload assessment has been recognized as one of the greatest improvement needs of the judicial system, but there is a lack of data needed to estimate the workload and correlate it properly with the resources. Hungary

General motives	Society - Organization	Organization - Individual
<i>Control</i>	<i>Managing resources and controlling performance deviations</i> • How to include different organizational circumstances and workload situations in balanced commensuration of goals and allocation of resources? • How to respect organizational independence by sensitively intervening in productivity and performance deviations?	<i>Detecting and controlling performance deviations</i> • How to balance between output, timeliness, and quality (“the trilemma”) in setting individual goals and allocating work? • How to utilize performance measurement data in detecting and sensitively intervening in productivity and performance deviations?
<i>Motivate</i>	<i>Encouraging court management to improve performance</i> • How to balance between output, timeliness, and quality in goal setting: the “trilemma” of setting performance targets? • How to balance between “punishment of performance” and “circle of impoverishment” in output-based funding practices?	<i>Encouraging judges to improve their performance</i> • How to create measurement practices which support independence, self-management, peer-control, and professional pride and decrease the need for controlling and intervening? • How to create opportunities for judges to participate in planning performance measurement processes and procedures?
<i>Learn</i>	<i>Improving and developing the entire judiciary in society</i> • The need for measures has been identified although existing practices are rare • How to overcome inherent organizational circumstances and strong organizational identities hampering benchmarking and distribution of good practices?	<i>Developing the individual courts and the judges</i> • The need for measures has been identified although existing practices are rare • How to include measures which support the exchange of knowledge and experiences effectively (e.g. different type of mentoring practices) and respect self-management and independence?

Table 2.
Summary of the main tasks related to output measurement systems in justice organizations

has implemented a system which aims to detect large workload differences, as they found out that there are courts which had to deal with over two times more incoming cases than other courts. In 2012, Hungary announced a plan for developing a scheme for classifying incoming cases based on their type and difficulty level. The scale of the scores ranges from 10 to 50. Also, for example, Finland, Sweden, and Estonia have established detailed weighted caseload systems for resource allocation purposes where all cases have weighted scores (e.g. Finland scores cases from 0.1 to 5.4 points). These types of systems have had good acceptance in courts, even though it is acknowledged that the appropriateness of the scores needs to be constantly updated based on practical experiences. It can be said that, due to the large variations in the time consumption of different cases, even a rough and approximate weighted caseload system is better than not weighting the cases at all.

Secondly, the control motive is related to the task of controlling the productivity and performance deviations of individual courts. The overall aim of the task is to keep the production processes of courts in “a normal stage” with a balanced pending case inventory and sufficient output levels. Due to the independent nature of the court organizations, the control efforts and possible intervening need to be carried out while respecting the autonomic nature of the court. Therefore, the control efforts aim to set general control limits with built-in early-warning mechanisms, instead of targeting the control efforts at the handling of individual cases or at the performance problems of individual professionals.

It became evident in the data analysis that issues related to throughput time and delayed cases are the easiest performance deviation area to control at society level. In time related issues, societies usually set limits and allow variations within the limits. For example, in France the goal is that 2/3 of the cases need to be handled in the set mandatory time limit. Sweden has implemented a sophisticated and detailed system for controlling performance deviations. In the Swedish system, a balanced pending case inventory refers to the number of pending cases which is “normal” and in balance with the resources and still allows a court to meet the set timeframes. In Sweden, the government has mandated that 75% of cases should be solved within the given time limits. The system uses the term “balanced inventory ratio” which refers to the ratio of a balanced pending case inventory and the number of incoming cases per year. The Swedish method is based on the empirical evidence that there is a clear linear relation between the throughput time and the inventory ratio.

Even though setting general time limits enable a sensitive approach to controlling performance deviations at court level, it contains challenges and has led to gaming behavior and perverse effects. Most of all, it has led to excessive solving of simpler cases to meet the limits and thus to long delays for the more complex cases. Within the control limits, it is possible that some cases get extremely delayed with no control mechanisms to detect them.

The motivate category at the society-organization measurement level includes tasks related to goal setting and output-based funding practices. Practices related to these areas have an impact on court management’s motivation to improve all aspects of an organization’s performance. The identified tasks in the motivate category are linked to the control tasks, as the sense of fairness and objectivity are the premises for the motivation to improve performance. The central motivation tasks in the data analysis were related to the so called “trilemma” of setting performance targets for courts. Courts have three important performance target areas: number of outputs, quality, and timeliness. In all analyzed countries, society level motivation efforts concentrate heavily on the improvement of the output number, creating a balancing problem between the accomplishment of all target areas. In many countries, for example in Hungary, recent reforms have put an even greater emphasis on the most visible numerical indicators which can only include certain aspects of the overall performance. This has led to a situation where society strongly motivates the accomplishment of output targets and the inherent values and ethics of professionals motivate the accomplishment of quality. Timeliness targets do not have equally clear motivational mechanisms, leading to delays and problems in the timely handling of all cases.

Excessive highlighting of the number of outputs also affects the motivational and rewarding aspects related to output-based funding. The inherent dilemma of providing financial incentives without punishment of performance or impoverishing the poorly performing organizations was relevant in all analyzed countries. If good output performance is rewarded by extra funding, it indirectly “punishes” the weaker organizations. Under market economy this would not be a problem, but in the case of public organizations providing public services, it creates problems of equal services to citizens. In the name of equality, even the poorly performing organizations need to be ensured adequate funding to deal with the workload. This sometimes leads to situations where extra funding is given to the underperformers. If good output performance is not rewarded, there are no incentives to raise performance above average.

In the Netherlands, the budget of a court fluctuates depending on the amount of output (95% of the court budget is based on the number of output). The more judicial cases are solved, the more budget will be assigned to the court. If a court produces more than expected, they receive 70% of the agreed price per product from the equalization account. When producing less than agreed, they must deposit 70% of the agreed prices of the cases not finalized into the equalization account. This is a strong stimulus for the courts to enhance production, but the system has been criticized. The system can also easily lead to gaming and

other unwanted behavior. Many Dutch judges have signed a manifest to protest the court financing system. The main critique is connected to the excess focus on output targets. The law provides that considerations of quality should play a role in determining the price of the case categories. It is argued, however, that the considerations of quality do not actually influence the financing of courts. The courts have responded to the problem by drafting professional standards and a system for the better integration of different targets into judicial budgeting.

The learn category measures related to the improvement and development of the entire judiciary are rare in the studied countries, but the need for them is clearly identified. The potential for using qualitative evaluation tools and performance data and measures to learn is recognized across judiciaries. One example is Italy, where it is recognized that the considerable amount of data created and managed by the courts has a social and economic impact (given that most disputes between people, companies, and public institutions find a resolution within the court). Making such data freely available to the public, organizations and institutes would improve knowledge of the many events affected or processed by the courts. There is an ongoing data-driven project aiming to make available the data collected by the justice system at national level. The action is a learning initiative based on data collection, analysis, and dissemination of the data on the Italian judicial system. The main objectives of the project are to support the measurement and benchmarking of the activities carried out by the judicial system and to enhance transparency and public accountability.

It also became evident that overemphasizing different organizational circumstances and strong organizational identities hamper benchmarking and distribution of good practices in the judicial system. Measurement does produce comparative data and would allow comparative learning and the detection of differences. In addition, a lot of feedback and assessment data is already collected. However, this measuring data is not currently utilized effectively for learning tasks and purposes. For example, in the improvement projects carried out in the Finnish judiciary, several means to improve the dissemination of the project results across court organizations were designed. The challenges in these systematic benchmarking efforts were connected to the emphasis on the differences in operating environments, circumstances, and case-structures between courts (so called “not-invented-here attitude”). Improvement in learning from other courts would require changes in the mindset related to organizational independence and autonomy connected to operational activities and improvement.

4.2 Tasks of output measurement at the organization-individual level

The control category at the organization-individual measurement level was found to include two tasks both aiming to detect and control the performance deviations of individual professionals. The challenge of balancing goal setting – “the trilemma” between output, timeliness, and quality goals – was clearly evident also at the organization-individual level. At this level, the tasks and effects of the trilemma relate more to control and work allocation practices than motivation (as at the society-organization level). The basic premise is that individual professionals are expected to produce the same number of outputs. Average throughput-time is also emphasized in controlling the performance of individual judges. These goals combined can cause perverse effects by encouraging excessive solving of simpler and newer cases to reach the output and throughput time goals. For example, in France the target that two-thirds of the cases be handled within a mandatory time limit has produced different type of unwanted behavior (e.g. a crowding-out effect on other cases for which the period of non-mandatory processing shall then be extended).

The weighted caseload systems used in different countries enable the comparison of output levels in more detail and reduce the gaming of the numbers. The weighted caseload

systems are also used to allocate cases between judges as evenly as possible. For example, in Hungary the system of weighting cases is seen to significantly improve the objectivity of goal setting and case allocation between judges. Due to the introduced case weights, the workload of the judges can be compared and balanced much more easily and more reliable data on the performance of individual judges can be obtained. The challenge recognized was the fact that the weight is assigned to the case at the very beginning of the trial (initial weight), and the complexity of the case can change later on during the process. As a response to the challenge, courts have introduced the system of “post-weighting” cases to obtain a more reliable calculation of individual workload and performance.

The other control task relates to situations where some individuals cannot reach their goals. When these performance deviations are detected, an intervention needs to be carried out while respecting the autonomic position of the professionals. Managers need to constantly balance between the need to intervene in performance problems and respecting autonomy. An important task of managers is to make use of the concrete support received from the set targets in detecting and sensitively intervening in productivity and performance deviations.

Different types of systems which proactively detect performance deviations have been implemented in many countries. These systems enable “early-warning” signals and produce transparent and real-time performance data for court management and self-management purposes. For example, in Slovenia there is a tool called “the President’s Dashboard” available to all court Presidents. The tool provides real-time information on individual performance. In Finland, a time-frame alarm system has been established to improve personal work planning, to reduce the number of pending cases and to eliminate long delays. The idea of the system is that those cases which are in danger to lag behind are detected early on when the set timeframe can still be reached. With the help of the time-frame alarm system, individual workers can monitor and control their own case inventory and schedule the work while taking account of old cases. The listings of pending cases are available and transparent to all in the organization. Also, “late warning” systems facilitate self-management. For example, in Estonia, at the beginning of each year, all judges get a list of “old” cases, and they need to provide explanations on why there is no final judgement yet. In every following quarter, the judges have to describe how the listed cases have proceeded since their reporting. Thanks to the system, the number of old cases has decreased by nearly 10 times. Also, for example Italy and Austria have similar checklist systems implemented.

The motivate category at the organization-individual measurement level also includes tasks which aim to support independence, professional pride, and self-management. Another important motivation task is to create opportunities for the professionals to influence the performance measures and participate in the target setting process and procedures. Setting unbiased targets and keeping the performance information completely transparent supports self-management and peer-control. Self-management, work ethics and professional pride are important means especially in ensuring the quality of cases. The identified tasks in the motivation category encourage judges to improve their own performance also in relation to productivity and timeliness, and thus reduce the need for controlling and intervening. It became evident in the analysis that, in recent years, the emphasis and importance of productivity as part of professional work has increased. The transparency of performance data also has a significant impact on the motivation to improve all aspects of performance.

An important task in strengthening self-management is to create opportunities for individual professionals to participate and influence target setting and performance measurement. In the data analysis, several performance improvement projects were identified where the systematic and wide participation of professionals significantly improved the results and success of implementation. For example, in Hungary, a timeliness – project (The “Debrecen model”) was carried out with a strong emphasis on the bottom-up

approach of improvement. The project targeted a comprehensive change in the attitudes of individual professionals with good results. In Finland, quality projects have been carried out in the jurisdiction of the Rovaniemi Court of Appeal, where the most central stakeholders of the jurisdiction have been actively involved. Based on the projects' results, it was concluded that a participatory, bottom-up approach increased peer-to-peer interaction between judges, improved attitudes towards change, encouraged discussion about work productivity among professionals, and created a culture for improvement.

Learn category measures, as well as goals and measurement practices which directly support learning aspects at the organization-individual level, are practically non-existent. However, the need for them has been widely recognized in the studies countries. In particular, there is a need to increase the understanding of individual professionals related to the overall performance situation of the organization and the individual's own role in creating productivity and well-performing processes (namely understanding how the "court-factory" works and functions). Also, in the learn category, the measuring tasks need to respect self-management and independency. Designing ways to exchange knowledge between peers could be beneficial – similarly to the participatory improvement projects carried out in different countries or creating different types of mentoring practices. Learning should be based on individual needs. Every judge has regular and mandatory judicial training based on their individually assessed needs. This type of system could be broadened to cover performance management related assessment and training (for example monitoring and work planning skills).

5. Discussion and conclusion

This study aimed to empirically explore and analyze the concrete tasks and inherent challenges of output measurement in European justice organizations. The overall goal was to include in the analysis the main general motives of performance measurement and the main stakeholders' perspectives. The strength and main source of originality of the study is that it provides a structured overview of output measurement practices and challenges in the professional field of justice. The study also highlights the importance of output measurement as a part of balanced performance measurement systems.

As the study use secondary source of evidence, the challenge is the lack of detailed descriptions related to output measurement practices. Therefore, the analysis required a relatively simplified analysis framework and to some extent subjective interpretation. The project reports deal with the years 2014–2017 and the results do not include any developments undertaken after that. As an addition, the analysis does not include all European countries.

The main implications of the study relate to the roles and motives of output measurement in professional organizations in general. The main implications point out that even though output measurement emphasizes control motive, there is room for development especially in self-management purposes and in learning from measurement information.

It can be concluded that there exists a clear need to study the possibilities to improve output measurement practices in professional public organizations, even though the research focus has shifted to the improvement of outcome measurement (see, e.g. [Barbato and Turri, 2017](#); [Jääskeläinen and Lönnqvist, 2011](#)). Achieving a flawless output measurement system at once is impossible, but a planned, systematic, and conscious improvement process is needed. Thus, the improvement of output measurement needs to be a continuing process, highlighting the importance of organizational and individual participation and their acceptance of the practices.

Previous studies have highlighted budgeting and resource allocation decisions and tasks in relation to output measurement in professional public organizations (e.g. [de Bruin, 2002](#);

Francesco and Guarini, 2018; Hur, 2018; Jääskeläinen and Roitto, 2014; Kristiansen, 2017). This study also confirmed that output measurement tasks, especially at the society-organization level, are largely driven by resource allocation needs. Furthermore, the need for objectivity and a sense of fairness proved to be central elements in designing and improving resource allocation practices and tasks. Objectivity and fairness should be considered fundamental parameters in designing budgeting and MBOR practices at society level, as well as work allocation practices at individual level. The analysis revealed that, in the European judicial sector, several systems and tools have been designed and implemented in order to improve the objectivity and impartiality of output measurement, budgeting, and resource allocation practices. The application of these types of systems should be further improved, even though it involves multidimensional and complex efforts connected to highly intangible outputs produced in a process with several uncontrollable circumstances. The study indicated that the sense of fairness and objectivity concerning the output measures have a large impact on both the approval of control tasks and to the success of motivational tasks.

In the judicial sector, the used output measures are clearly designed for mainly control purposes and designed based on practical reasons that are too straightforward. It has been argued that this is quite common in professional public organizations and can lead to simplifications of reality and unwanted behavior (e.g. de Bruijn, 2011; Johansson, 2015). The main indicators used in the justice system (number of output and average throughput time) have several practical advantages. They are reliable and unbiased, easy to quantify and understand, as well as functional in controlling performance deviations. It can also be said that the use of these types of simplifying measures enables society and managers to intervene in performance deviation problems sensitively and objectively. However, the analysis clearly showed that simplifying measures does not have good validity, leading to different types of gaming and unwanted behavior. In the studied countries this has caused resistance towards measurement practices and long delays for some judicial cases. It can also be argued that the measures are not informative enough to be used for motivation and learning tasks and purposes, highlighting the control tasks of output measurement even further.

The study focused attention on the need for designing measurement practices which support the self-management of professionals. Self-management proved to play a central role in the motivational aspects of output measurement in the judicial sector and also decreased the need for controlling. Based on the study, self-management reduces possible tension between managers and professionals and decreases the number of professionals rejecting or abusing the output measures (see, e.g. de Bruijn, 2011). Self-management practices also increased the individuals' understanding of their role and contribution to the overall performance of the organization. Transparency of performance data and peer-control proved to play an important role in facilitating self-management. Previous studies highlighting the importance of giving individuals wide possibilities to influence and commit to performance measurement practices (e.g. Jääskeläinen and Roitto, 2014; Van Thiel and Leeuw, 2002) have touched on the concept of self-management. However, it can be concluded that self-management and ways of improving intrinsic motivation should be studied more in relation to performance measurement and management in professional public organizations.

The need to improve organizational learning from measurement information has been recognized in the literature (e.g. Kalgin *et al.*, 2018; Laihonen and Mäntylä, 2017). Similarly, the need for improving the learning tasks of output measurement became clear and evident across the studied judiciaries. The used indicators allow comparison in terms of objectivity and use of resources, but they are not sufficiently exploited in information sharing, benchmarking, and learning. Based on the study, the requirements for using measurement structures and procedures to learn will be emphasized even more in the future. This is why it can be concluded that improving the learning tasks of output measurement in professional public organizations is a central area in need of future research and improvement efforts.

References

- Barbato, G. and Turri, M. (2017), "Understanding public performance measurement through theoretical pluralism", *International Journal of Public Sector Management*, Vol. 30 No. 1, pp. 15-30.
- Behn, R.D. (2003), "Why measure performance? Different purposes require different measures", *Public Administration Review*, Vol. 35 No. 5, pp. 586-606.
- Björk, L., Szucs, S. and Härenstam, A. (2014), "Measuring capacity to perform across local government services– managers' perceptions", *International Journal of Public Sector Management*, Vol. 27 No. 1, pp. 26-38.
- Boland, T. and Fowler, A. (2000), "A systems perspective of performance management in public sector organisations", *International Journal of Public Sector Management*, Vol. 13 No. 5, pp. 417-446.
- Charbonneau, E., Bromberg, D.E. and Henderson, A.C. (2015), "Performance improvement, culture, and regimes. Evidence from the Ontario Municipal performance measurement program, 2000-2012", *International Journal of Public Sector Management*, Vol. 28 No. 2, pp. 105-120.
- Contini, F., Anley, S. and Rottman, D.B. (2014), "Evaluating judicial performance: editors' introduction", *Onati Socio-Legal Series*, Vol. 4 No. 5, pp. 836-842.
- de Bruijn, H. (2002), *Managing Performance in the Public Sector*, Routledge, New York, p. 129, ISBN: 0-415-30037-1.
- de Bruijn, H. (2011), *Managing Professionals*, Routledge, New York, p. 195, ISBN: 978-0-415-56508-0.
- De Santis, L. and Emery, Y. (2017), "Comparing stakeholders' expectations of "good justice" in Switzerland: toward a 'hybrid' judicial culture", *Justice System Journal*, Vol. 38 No. 1, pp. 78-111.
- Dobija, D., Górska, A.M., Grossi, G. and Strzelczyk, W. (2019), "Rational and symbolic uses of performance measurement - experiences from Polish universities", *Accounting, Auditing and Accountability Journal*, Vol. 32 No. 3, pp. 750-781.
- Elg, M., Broryd, K.P. and Kollberg, B. (2013), "Performance measurement to drive improvements in healthcare practice", *International Journal of Operations and Production Management*, Vol. 33 Nos 11/12, pp. 1623-1661.
- Førsund, F.R. (2012), "Measuring efficiency in the public sector", Memorandum 09/2012, Oslo University, Department of Economics, p. 29.
- Francesconi, A. and Guarini, E. (2018), "Performance-based funding and internal resource allocation: the case of Italian universities", in Borgonovi, E., Anessi-Pessina, E. and Bianchi, C. (Eds), *Outcome-Based Performance Management in the Public Sector, System Dynamics for Performance Management 2*, Springer International Publishing AG, pp. 289-306.
- Garlatti, A., Fedele, P. and Ianniello, M. (2018), "The iron law of unintended effects, again? Outcome measures and blame-avoidance", in Borgonovi, E., Anessi-Pessina, E. and Bianchi, C. (Eds), *Outcome-Based Performance Management in the Public Sector, System Dynamics for Performance Management 2*, Springer International Publishing AG, pp. 45-62.
- Goh, S.C., Elliott, C. and Richards, G. (2015), "Performance management in Canadian public organizations: findings of a multi-case study", *International Journal of Productivity and Performance Management*, Vol. 64 No. 2, pp. 157-174.
- Hood, C. (1991), "A public management for all seasons?", *Public Administration*, Spring, Vol. 69, pp. 3-19.
- Hur, H. (2018), "The unintended consequences of a pay-for-performance rule change. U.S. Department of Defense's National Security Personnel System", *International Journal of Public Sector Management*, Vol. 31 No. 6, pp. 654-671.
- Ittner, C. and Larcker, D. (2002), "Determinants of performance measure choices in worker incentive plans", *Journal of Labor Economics*, Vol. 20 No. 2, pp. 568-590.
- Jääskeläinen, A. and Lönnqvist, A. (2011), "Public service productivity: how to capture outputs?", *International Journal of Public Sector Management*, Vol. 24 No. 4, pp. 289-302.

- Jääskeläinen, A. and Raito, J.-M. (2014), "Drivers of personnel satisfaction towards performance information usage: findings from Finnish municipalities", *International Journal of Public Sector Management*, Vol. 27 No. 6, pp. 530-547.
- Jääskeläinen, A., Laiho, H. and Lönnqvist, A. (2014), "Distinctive features of service performance measurement", *International Journal of Operations and Production Management*, Vol. 34 No. 12, pp. 1466-1486.
- Johansson, V. (2015), "When will we ever learn?", *The NISPAcee Journal of Public Administration and Policy*, Vol. 8 No. 2, Winter 2015/2016 pp. 149-170.
- Kalgin, A., Podolskiy, D., Parfenteva, D. and Campbell, J.W. (2018), "Performance management and job-goal alignment. A conditional process model of turnover intention in the public sector", *International Journal of Public Sector Management*, Vol. 31 No. 1, pp. 65-80.
- Kerpershoek, E., Groenleer, M. and de Bruijn, H. (2016), "Unintended responses to performance management in Dutch hospital care", *Public Management Review*, Vol. 18 No. 3, pp. 417-436.
- Kristiansen, M.B. (2017), "Performance contracting. Understanding variation across agencies in Danish Central Government", *International Journal of Public Sector Management*, Vol. 30 No. 2, pp. 170-188.
- Laiho, H. and Mäntylä, S. (2017), "Principles of performance dialogue in public administration", *International Journal of Public Sector Management*, Vol. 30 No. 5, pp. 414-428.
- Lienhard, A., Kettiger, D. and Winkler, D. (2012), "Status of court management in Switzerland", *International Journal for Court Administration*, Vol. 4 No. 3, pp. 1-27.
- Linna, P., Pekkola, S., Ukko, J. and Melkas, H. (2010), "Defining and measuring productivity in the public sector: managerial perceptions", *International Journal of Public Sector Management*, Vol. 23 No. 3, pp. 300-320.
- Lipe, M. and Salterio, S. (2000), "The balanced scorecard: judgmental effects of common and unique performance measures", *Accounting Review*, Vol. 75, pp. 283-298.
- Lowe, T. (2013), "New development: the paradox of outcomes—the more we measure, the less we understand", *Public Money and Management*, Vol. 33 No. 3, pp. 213-216.
- Malina, M. and Selto, F. (2004), "Choice and change of measures in performance measurement models", *Management Accounting Research*, Vol. 15, pp. 441-469.
- Mandl, U., Dierx, A. and Ilzkovitz, F. (2008), "The effectiveness and efficiency of public spending", *European Economy - Economic Papers, 2008*, Directorate General Economic and Financial Affairs (DG ECFIN), European Commission, Vol. 301, pp. 1-34, ISBN: 978-92-79-08226-9.
- McGeough, F. (2015), "Performance reporting in Ireland: the ongoing gap between rhetoric and reality", *International Journal of Public Sector Management*, Vol. 28 No. 1, pp. 2-10.
- McPhee, I. (2005), "Outcomes and outputs: are we managing better as a result?", *CPA National Public Sector Convention*, Friday 20 May 2005.
- Merryman, J.H. and Pérez-Perdomo, R. (2007), *The Civil Law Tradition: An Introduction to the Legal Systems of Europe and Latin America*, 3rd ed., Stanford University Press, Stanford, CA.
- Pekkanen, P. (2011), "Delay reduction in courts of justice – possibilities and challenges of process improvement in professional public organizations", *Acta Universitatis Lappeenrantaensis, Lappeenranta University of Technology*, Vol. 444, p. 200, ISBN: 978-952-265-147-1.
- Pollitt, C. (2006), "Performance management in practice: a comparative study of executive agencies", *Journal of Public Administration Research and Theory*, Vol. 16 No. 1, pp. 25-44.
- Pollitt, C. and Bouckaert, G. (2000), *Public Management Reform: A Comparative Analysis*, Oxford University Press, USA, ISBN: 0-19-829596-0.
- Radnor, Z. (2008), "Muddled, massaging, manoeuvring or manipulated? A typology of organizational gaming", *International Journal of Productivity and Performance Management*, Vol. 57 No. 4, pp. 316-328.

-
- Radnor, Z. and McGuire, M. (2004), "Performance management in the public sector: fact of fiction?", *International Journal of Productivity and Performance Management*, Vol. 53 Nos 3/4, pp. 245-260.
- Rajala, T., Laihonon, H. and Vakkuri, J. (2018), "Shifting from output to outcome measurement in public administration - arguments revisited", in Borgonovi, E., Anessi-Pessina, E. and Bianchi, C. (Eds), *Outcome-Based Performance Management in the Public Sector, System Dynamics for Performance Management 2*, Springer International Publishing AG, pp. 3-23.
- Rantanen, H., Kulmala, H.I., Lönnqvist, A. and Kujansivu, P. (2007), "Performance measurement systems in the Finnish public sector", *International Journal of Public Sector Management*, Vol. 20 No. 5, pp. 415-433.
- Speklé, R.F. and Verbeeten, F.H.M. (2014), "The use of performance measurement systems in the public sector: effects on performance", *Management Accounting Research*, Vol. 25 No. 2, pp. 131-146.
- Tabi, M.T. (2013), "Positioning the performance of public service. When management tools are in conflict: the case of urban water management", *International Journal of Public Sector Performance Management*, Vol. 2 No. 2, pp. 175-198.
- Van Dooren, W., Bouckaert, G. and Halligan, J. (2015), *Performance Management in the Public Sector*, 2nd ed., Routledge, London, ISBN: 9781315817590.
- Van Thiel, S. and Leeuw, F.L. (2002), "The performance paradox in the public sector", *Public Performance and Management Review*, Vol. 25 No. 3, pp. 267-281.
- Vecchi, G. (2018), "Performance management uses, outcome measures and mechanisms-based explanations. The case of the judicial sector", in Borgonovi, E., Anessi-Pessina, E. and Bianchi, C. (Eds), *Outcome-Based Performance Management in the Public Sector, System Dynamics for Performance Management 2*, Springer International Publishing AG, Switzerland, pp. 331-348.
- Vogel, R. (2013), "Tracing the quest for performance in Germany's public sector: institutional entrepreneurship with rhetorical strategies of persuasion", *International Journal of Public Sector Performance Management*, Vol. 2 No. 2, pp. 123-140.

About the authors

Petra Pekkanen Associate Professor of Supply Chain and Operations Management at LUT University. Doctor of Science in Technology (D.Sc.tech.) in the field of Industrial Engineering and Management. Main research interest is process improvement and operations management in professional services and organizations. Petra Pekkanen is the corresponding author and can be contacted at: petra.pekkanen@lut.fi

Timo Pirttilä Professor of Supply Chain and Operations Management at LUT University. Main research interest is process improvement and operations management in professional services and organizations.

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com