

System Dynamics for Performance Management 2

Elio Borgonovi
Eugenio Anessi-Pessina
Carmine Bianchi *Editors*

Outcome-Based Performance Management in the Public Sector

System Dynamics for Performance Management

Volume 2

Series Editor

Carmine Bianchi, CED4-System Dynamics Group, University of Palermo,
Palermo, Italy

More information about this series at <http://www.springer.com/series/13452>

Scientific Committee for System Dynamics for Performance Management

- Luca Anselmi** University of Pisa, Italy—Professor of Public Administration
- David Birdsall** Baruch College/CUNY, USA—Dean, School of Public Affairs
- Elio Borgonovi** Bocconi University, Milan, Italy—Professor of Economics and Management of Public Administration
- Tony Bovaird** University of Birmingham, UK—Professor of Public Management and Policy
- Dario Cavenago** Bicocca University, Milan, Italy—Professor of Public Management
- Lino Cinquini** Scuola Superiore Sant'Anna, Pisa, Italy—Professor of Business Administration
- Paal I. Davidsen** University of Bergen, Norway—Professor of System Dynamics, Chair of the System Dynamics Group
- David Lane** Henley Business School, UK—Professor of Informatics
- Manuel London** State University of New York at Stony Brook, USA—Distinguished Professor of Management
- Luciano Marchi** University of Pisa, Italy—Professor of Planning & Control Systems
- Marco Meneguzzo** Università della Svizzera Italiana, Lugano—Switzerland; University Tor Vergata, Rome, Italy—Professor of Public Management
- Riccardo Mussari** University of Siena, Italy—Professor of Public Management
- Ozge Pala** Koç University, Istanbul, Turkey—Assistant Professor of Management Sciences
- Guy Peters** University of Pittsburgh, USA—Maurice Falk Professor, Department of Political Science
- Angelo Riccaboni** University of Siena, Italy—Professor of Planning & Control Systems
- William C. Rivenbark** University of North Carolina at Chapel Hill, USA—MPA Program Director, School of Government
- Etienne Rouwette** Nijmegen School of Management, The Netherlands—Associate Professor of Research Methodology and System Dynamics
- Salvatore Rotella** Riverside College, California, USA—Chancellor Emeritus and Professor of Political Science
- Khalid Saeed** Worcester Polytechnic Institute, USA—Professor of System Dynamics
- Markus Schwaninger** University of St Gallen, Switzerland—Professor of Management
- Carlo Sorci** University of Palermo, Italy—Professor of Business Management
- Jürgen Strohhecker** Frankfurt School of Finance & Management, Germany—Professor for Business Administration, Operations and Cost Management
- Jarmo Vakkuri** University of Tampere, Finland—Professor of Local Government Accounting & Finance
- Wouter Van Dooren** University of Antwerp, Belgium—Associate Professor of Public Management
- David Wheat** University of Bergen, Norway—Associate Professor in System Dynamics

Elio Borgonovi · Eugenio Anessi-Pessina
Carmine Bianchi
Editors

Outcome-Based Performance Management in the Public Sector

CED⁴
System Dynamics Group

 Springer

Editors

Elio Borgonovi
Department of Policy Analysis and Public
Management
Bocconi University
Milan
Italy

Carmine Bianchi
CED4-System Dynamics Group
University of Palermo
Palermo
Italy

Eugenio Anessi-Pessina
Department of Economics and Business
Management Science
Catholic University of the Sacred Heart
Milan
Italy

ISSN 2367-0940 ISSN 2367-0959 (electronic)
System Dynamics for Performance Management
ISBN 978-3-319-57017-4 ISBN 978-3-319-57018-1 (eBook)
DOI 10.1007/978-3-319-57018-1

Library of Congress Control Number: 2017939624

© Springer International Publishing AG 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Part I The Design of Outcome-Based Performance Management Systems in the Public Sector

- 1 Shifting from Output to Outcome Measurement in Public Administration-Arguments Revisited** 3
Tomi Rajala, Harri Laihonen and Jarmo Vakkuri
- 2 Organizational Performance in the Italian Health care Sector** 25
Alessandro Spano and Anna Aroni
- 3 The Iron Law of Unintended Effects, Again? Outcome Measures and Blame-Avoidance** 45
Andrea Garlatti, Paolo Fedele and Mario Ianniello
- 4 Designing Outcome-Based Performance Management Systems to Assess Policies Impacting on Caesarean Section Rate: An Analysis of the Sicilian Maternity Pathway** 63
Enzo Bivona and Federico Cosenz

Part II Towards Outcome-Based Performance Management: Experiences and Trends from Different Countries

- 5 The Scottish Government's System of Outcome-Based Performance Management: A Case Study of the National Performance Framework and Scotland Performs** 81
Bobby Mackie
- 6 Governance Structures and the "(De)Politicization" of Performance Measures** 107
Henrik P. Minassians and Ravi K. Roy
- 7 Italian Public Administration Reform: What are the Limits of Financial Performance Measures?** 121
Paolo Ricci and Renato Civitillo

Part III Enhancing Governance and Inter-institutional Coordination through Outcome-Based Performance Management	
8 Measuring Coordination and Coherence: Assessing Performance Across the Public Sector	143
Carmine Bianchi and B.G. Peters	
9 Outcome-Based Performance Management in the Public Sector: What Role for Inter-organizational ICT Networks?	161
Luca Brusati, Paolo Fedele, Mario Ianniello and Silvia Iacuzzi	
10 Disappointing Outcomes: Can Implementation Modeling Help?	179
I. David Wheat and Eugene Bardach	
Part IV Applying Outcome-Based Performance Management: Experiences from Different Public Sector Domains	
11 Evaluation of Innovation Performance in the Public Sector: A Systematic Review of Studies	203
Greta Nasi, Maria Cucciniello and Virginia Degara	
12 Measuring the Impact of Training on Public Manager Performance: The Case of Canton Ticino (An Ex-Post Analysis)	225
Andrea Martone, Filippo Sciaroni and Alan Righetti	
13 Analysing Corruption: Effects on the Transparency of Public Administrations	251
Isabella Fadda, Paola Paglietti, Elisabetta Reginato and Aldo Pavan	
14 Performance Measurement Systems in Universities: A Critical Review of the Italian System	269
Natalia Aversano, Francesca Manes-Rossi and Paolo Tartaglia-Polcini	
15 Performance-Based Funding and Internal Resource Allocation: The Case of Italian Universities	289
Andrea Francesconi and Enrico Guarini	
16 Performance Reporting in Italian Public Universities: Activities in Support of Research, Teaching and the “Third Mission”	307
Elisa Bonollo and Mara Zuccardi Merli	
17 Performance Management Uses, Outcome Measures and Mechanisms-Based Explanations. The Case of the Judicial Sector	331
Giancarlo Vecchi	

18 Performance Management and Evaluation of Large-Scale Events in a Multistakeholder Engagement Perspective: The Case of the Extraordinary Jubilee of Mercy 349
Marco Meneguzzo, Gloria Fiorani and Rocco Frondizi

Part V Dealing with the Challenges of Outcome-Based Performance Management: Innovative Methods and Tools

19 Performance Benchmarking of School Districts in New York State 373
Thomas R. Sexton, Christie Comunale, Michael Shane Higuera and Kelly Stickle

20 Efficiency Versus Effectiveness in Hospitals: A Dynamic Simulation Approach 397
Markus Schwaninger and Johann Klocker

21 Public Policy Design for Climate Change Adaptation: A Dynamic Performance Management Approach to Enhance Resilience 425
Hugo Herrera

Index 447

Introduction

One of the pillars of management studies, theories, and practice is that “nothing can be improved if it is not measured.” This statement is almost unanimously accepted, notwithstanding Einstein’s words whereby “not everything that counts can be counted, and not everything that can be counted counts.” Anyway, we think that most people would agree on two main concepts:

- The measurement of individual (microlevel), organizational (mesolevel), and system (macrolevel) performance is a strong leverage to pursue improvements and activate change processes.
- People’s motivation is another relevant leverage, even when it cannot be measured, as demonstrated by behavioral economics and management and by intrinsic motivation theory (Perry 1990, 2010).

It is thus necessary to answer a fundamental question: What is performance? From a theoretical point of view, the answer is fairly simple for private enterprises that operate in a market arena. In principle, performance coincides with profit and shareholder remuneration in the classic model of enterprise and with stakeholder rewards (Freeman 1984) in the modern concept of enterprise related to CSR and CSV (Porter and Kramer 2011) theories. From a practical point of view, however, accountants know very well the difficulties involved in measuring profit and stakeholder rewards, as reflected by the many different methodological techniques and assumptions that have been developed in different countries and different institutional environments. International Financial Reporting Standards (IFRS) attempt to overcome these differences by comparing enterprise performances in the global economy. Nevertheless, many technical issues remain unsolved, as shown by the thousands of books and papers on this matter and by the hundreds (perhaps thousands) of official documents aimed at applying IFRS principles, criteria, and rules in various countries. In the private sector, other performance indicators to measure, analyze, and evaluate enterprises’ competitiveness and success include market share, number of clients, customer satisfaction, quality of goods and

services, productivity, cost of products (total and itemized), turnover (total and broken down by products and markets), and categories of products.

In public administration, there is far less agreement on the concept of performance. There are, of course, financial performance indicators such as surplus and deficit, amounts of revenues and expenditures, revenues-to-expenditure ratios, capital-to-current expenditure ratios under traditional budgetary accounting as well as profit or loss, income and cost, income-to-cost ratios, investment-to-expense ratios under accrual accounting. Moreover, also in the public sector, there has recently been a movement toward the adoption of International Public Sector Accounting Standards (IPSAS). However, financial performance indicators are differently interpreted as reflecting success or failure under different political, institutional, socioeconomic situations.

Public sector performance, moreover, must largely be viewed as non-financial. Non-financial performance, however, has been defined differently under different models of public administration. In the traditional, “formal rights” model, which can be viewed as the first stage of the modern State, the dominant principle for the evaluation of public administration was compliance with laws and regulations. Therefore, the main implementation criteria were standardization and stabilization, while performance indicators were related to inputs control (budget allocation and actual use of resources) and conformance with standardized procedures.

In the second stage of the modern State, the so-called redistribution welfare model, the critical principles became the quality of policies (taxation and redistribution) and of priority setting among different groups of beneficiaries (individuals, families, disadvantaged groups, etc.). Thus, performance was expressed in terms of the amount of resources collected and allocated (again, input-related performances) and actual numbers of beneficiaries reached (often operationalized as the ratio between the pursued objectives of policies and the actual results of implementation).

In the third stage, which can be labeled as the “welfare of services” model, the output concept became more and more relevant. The implementation of policies increasingly required the physical combination and transformation of inputs into outputs as opposed to the simple transfer of financial resources from the government to groups of beneficiaries. Consequently, the quantity and quality of outputs as well as the government’s internal efficiency became more and more relevant, in line with the New Public Management approach. Consistently, performance systems paid increasing attention to the quantity and quality of outputs as well as to outputs–inputs ratios (productivity and efficiency of different resources such as personnel, equipment, and data). In the 1970s and 1980s, hundreds of indicators were developed for the whole organization and for specific government activities (e.g., garbage collection, social services, education, health, maintenance of public buildings, street construction and repair). These indicators were used for trend analyses within the same government organization (improvement or worsening of performances over time) as well as for comparisons and benchmarking across different administrations.

What is currently the last evolution of the modern State emerged in the late 1990s and was characterized by a shift from government to governance. Due to technological innovations, social changes, and public finance constraints, the new focus is on the external effects of public administration activities, that is, on outcomes and policy impacts. The new key assessment criteria are appropriateness and effectiveness of public administration. Consistently, performance measurement systems have evolved toward outcome measurement, outcome–output ratios, outcome–input ratios, policy impact indicators, and citizen satisfaction indicators. However, the adoption and implementation of these performance indicators has been problematic because often there is no agreement on what links outcomes with inputs.

In conclusion, performance measurement systems and issues in the public sector can thus be analyzed and interpreted in light of the following shifts:

- from an internal (input–output) to an external (output–outcome) perspective;
- from competition to collaboration or collaborative competition;
- from the separation between policymaking and administration/management to the integration of the policy cycle (co-analysis, co-decision, co-design, co-evaluation);
- from a focus on specialization to one on interdependencies;
- from silos responsibilities/financing to unitary responsibility/financing;
- from organizational unit performance to institutional performance;
- from professionals/bureaucrats to managers; and
- from economic to social evaluation.

This book is a collection of papers that discuss the more recent aspects of this evolution. It consists of five parts and includes 21 chapters. All of them combine a conceptual as well as an empirical approach. Although they all contribute to a systematic analysis of the topic from a theoretical standpoint, most of them also provide relevant insights on the practical experiences of shifting the performance management systems' paradigm from outputs to outcomes in the public sector. The field studies included in this book—under the shape of empirical cases, interviews, and data analysis—are related to the experiences developed in different countries.

Part I of this volume aims at shedding light on problems and issues implied in the design and implementation of “outcome-based” performance management systems in the public sector.

The main debating points that this part addresses are as follows:

- What arguments encourage politicians and public managers to stick with outputs while ignoring outcomes in performance measurement? What reasons could urge them to adopt also outcome measures?
- How to define organizational performance, with a particular focus on outcomes? How to measure it? How to design performance management systems that may go beyond the measurement of individual performance?
- Are there any unintended behavioral effects associated with the use of outcome performance measures? When designing and using outcome measures, is it

possible to predict the possibility that performance paradoxes will arise, i.e., that behavioral reactions of decision makers will only formally pursue the achievement of the outcomes for which they are made accountable?

- How to design performance management systems that may support decision makers in identifying and measuring the drivers impacting on outcomes? How can performance management support elected officials and administrators to adopt decisions that may impact on such drivers and therefore on outcomes?

Such debating points provide the core of the first four manuscripts hosted in this volume.

The book begins with a chapter by Tomi Rajala, Harri Laihonen, and Jarmo Vakkuri. The authors address the first set of the previously mentioned debating points. In order to discuss the arguments against, or in favor of the use of outcome performance measures, they propose a conceptual framework including five categories: information, controllability, legitimacy, nature of outcomes, and political conflict.

The second chapter, by Alessandro Spano and Anna Aroni, contributes to the framing of the second set of debating points, with a specific focus on health care in Italy. Their research, based on an in-depth analysis of the content of the documents published by a group of Italian public healthcare organizations, provides evidence of a significant variance in the way organizational performance is defined and measured. This difference is symptomatic of a difficulty deriving from the attempt to implement a top-down performance management system enforced by law—as it is in the case of Italy, not only in health care but in the entire public sector. This phenomenon is also a strong sign of how cultural issues together with professional and institutional factors systematically shape and affect the paradigm shift toward outcome-based performance management systems in the public sector.

The chapter by Andrea Garlatti, Paolo Fedele, and Mario Ianniello provides insights for the debate of the third group of questions. To this end, the authors analyze the case of the labor policies in an Italian region (Lombardy), where outcome performance measures have been adopted to foster decision makers' accountability. The field analysis suggests that although the policy had been successful, gaming behaviors by service providers worsened placement results.

The fourth set of questions is framed in the manuscript by Enzo Bivona and Federico Cosenz, with a specific focus on health care. In this regard, the authors analyze the case of Cesarean sections in an Italian region (Sicily). The causes behind the difficulties of the regional healthcare system to reduce the rate of such surgeries are discussed. The need of a dynamic performance management (DPM) system is advocated in order to enable decision makers to effectively pursue the desired outcomes. To describe the benefits of such approach for outcome-based performance management in the analyzed case, a conceptual DPM model is outlined and discussed by the authors.

Part II of the book illustrates the experiences, problems, and evolving trends in three different countries (Scotland, USA, and Italy) toward the adoption of

outcome-based performance management systems in the public sector. Such analyses are conducted at both the national and local government levels.

This part begins with a chapter by Bobby Mackie. The chapter illustrates the difficulties in accommodating outcomes in performance management systems in the Scottish national government. The author emphasizes how the analyzed case study suggests that it is possible to gradually overcome such difficulties by developing and encouraging an outcome-focused culture in public service provision. The chapter also remarks how the alignment between national and local government reforms is crucial in effectively implementing a paradigm shift from output to outcome-based performance management.

The second chapter in Part II, by Henrik Minassians and Ravi Roy, discusses the problem of the lack of consistency between performance measures and the overall strategic planning goals. Such problem is discussed in relation to the characteristics of: (a) the level of coordination between agencies in local government, (b) the features of politico-administrative systems, and (c) leadership style. To debate such issues, the chapter focuses on two research questions: (1) What role does the politico-administrative structure of local county governments play in the design of performance measures? and (2) How do elected officials use performance measures in their decision-making processes? To address such questions, the cases of two counties in Southern California are illustrated.

The third chapter in Part II, by Paolo Ricci and Renato Civitillo, addresses the intrinsic limitations of a performance management system that focuses only on financial measures. This problem is discussed under the perspective of the Italian public sector system, where the role of legislation in adopting performance management systems has been stronger than the perception of the need to develop other attributes as well, such as professional skills and an outcome-oriented performance culture.

Part III of this book frames how outcome-based performance management can enhance public governance and inter-institutional coordination. Often, governance and coordination are conceived as targets to pursue by only acting on legislative and administrative rules, and—more generally—on the institutional system design. Though such levers undoubtedly matter to enhance coordination and governance, they are not sufficient to this end. In fact, both professional/managerial and cultural factors are relevant in designing and implementing sustainable reforms that may insure effective new public governance.

The chapter by Carmine Bianchi and Guy Peters discusses the advantages of designing dynamic and outcome-based performance management systems to measure and foster inter-agency coordination in public service delivery. The cases of social/health and food policies, with a specific focus on the US system, are discussed. A generic dynamic performance management model to foster policy integration and service delivery in highly dynamic and complex systems is then illustrated and applied to food policies.

The chapter by Luca Brusati, Paolo Fedele, Mario Ianniello, and Silvia Iacuzzi explores how inter-organizational Information and Communication Technology (ICT) networks can improve performance in local economic development. To this

end, the case of the Friuli Venezia Giulia Autonomous Region in Italy is discussed. The case study demonstrates the powerful role of ICT in fostering inter-agency coordination and outcome-based performance management through better governance. The authors also remark that “silos thinking” can be effectively challenged if also a cultural change is implemented in the mental models of the decision makers.

The third chapter in Part III, by David Wheat and Eugene Bardach, faces the problem of policy coordination under a different viewpoint, with respect to the previous chapters. The authors discuss the causes of disappointing outcomes in policy implementation. They link such phenomenon to a lack of communication between actors in policy design and implementation. To counteract this problem, they propose the use of system dynamics modeling and simulation. This is more than just a simulation technique, since it provides a methodology that—through mapping and model-building facilitation—may enhance a better understanding of the relevant system structure and behavior. Therefore, it may support the design of more consistent and “robust” public policies, whose implementation can be conceptualized when policy design occurs. To illustrate how system dynamics modeling can be helpful to this end, the authors discuss a rather famous case in public administration literature: the so-called Oakland fiasco—analyzed by Pressman and Wildavsky—a project to combat persistent unemployment among minorities in Oakland, California, in the late 1960s.

Part IV of this book deals with the illustration of challenges and results from different public sector domains.

We begin with a chapter by Maria Cuccinello, Greta Nasi, and Virginia Degara. Through a systematic review of the literature, the authors discuss the status and trends in measuring the outcomes from innovation processes in the public sector. This is a longitudinal topic to many public sector fields. It also has a governance dimension, since the impact of innovation often encompasses the domains of different agencies, institutions, and stakeholders in public service delivery. Furthermore, it also crosses vertically two often-disconnected decision-making areas, i.e., policymaking and administration. Through the analysis of concrete experiences illustrated by the literature, the authors discuss what steps ahead should be made to move forward in the use of more outcome-oriented measures in evaluating the impact of innovation. This implies—among other things—an effort aimed at the following: (1) to better focus the specific features of innovation processes in the public sector and (2) to frame innovation processes and gauge their own outcomes in a more systematic and dynamic manner.

The next chapter, by Andrea Martone, Filippo Sciaroni, and Alan Righetti, debates another longitudinal topic, i.e., measuring the impact of training on the performance of public managers. The case of the Swiss Canton Ticino is illustrated.

The chapter by Isabella Fadda, Paola Paglietti, Elisabetta Reginato, and Aldo Pavan discusses the cause-and-effect relationship between corruption and transparency. Though such nexus may appear ambiguous, based on an empirical research on published reports by the Italian regions, the authors illustrate how corruption is a main cause of low transparency in reporting to various stakeholders.

The three subsequent chapters focus on the topic of outcome-based performance management in the domain of Italian Universities.

The chapter by Natalia Aversano, Francesca Manes Rossi, and Paolo Tartaglia Polcini illustrates and debates the development of performance measurement systems in Italian Universities and discusses possible strategies to adapt such systems toward international harmonization.

The chapter by Andrea Francesconi and Enrico Guarini discusses whether performance management systems of Italian Universities are able to gauge the measures on which such institutions are ranked and receive funding from the national Ministry of Education. Based on a field study on the reporting systems in Italian Universities, the authors remark how the quality of performance management in such institutions depends on their capability to gauge and keep under control the performance measures upon which the Italian Ministry of Education allocates university funding. The extent to which outcome-based performance management systems should be focused mainly on the measures used for external performance benchmarking or should also reflect also the specific internal context and strategy of an organization is a debated issue in both theory and practice.

The chapter by Elisa Bonollo and Mara Zuccardi Merli illustrates how Italian Public Universities have outlined their performance reporting on research, teaching, and the so-called ‘third mission’.

Giancarlo Vecchi, in another chapter, frames the topic of outcome-based performance management in the context of the Italian judicial sector. To this end, two cases are analyzed, i.e., the Court of Milan and the Public Prosecutor’s Office of Milan.

The concluding chapter in Part IV, by Marco Meneguzzo, Gloria Fiorani, and Rocco Frondizi, proposes a multidisciplinary approach to analyze the case of the Extraordinary Jubilee of Mercy. In this regard, they discuss the outcomes of such event, in terms of inter-institutional and collaborative governance, cross-sector collaboration, and joined-up government.

Part V of this book focuses on innovative methods and tools that may support decision makers in dealing with the challenges of outcome-based performance management in the public sector.

In particular, the chapter by Thomas Sexton, Christie Comunale, Michael Shane Higuera, and Kelly Stickle shows the advantages of Data Envelopment Analysis to enhance performance benchmarking, according to an outcome-based view. The case of New York State school districts is illustrated.

The two concluding chapters, by Markus Schwaninger and Johann Klocker, and by Hugo Herrera, illustrate the advantages of system dynamics modeling and simulation to enhance outcome-based performance management.

The former is focused on the analysis and discussion of an Austrian hospital case and, in particular, of its oncology section. The latter illustrates how system dynamics may foster outcome-based performance management to enhance resilience to climate change.

We hope that this book will shed light on a topic that still today demands for a deeper analysis (also through comparative research) and more empirical studies illustrating good and bad practices in the field.

To conclude this work, the editors wish to thank both the authors and the anonymous referees, without whom the publication of this volume might not have been possible.

References

Freeman, R. E. (1984). *Strategic management: A stakeholder approach*. Cambridge: Cambridge University Press.

Perry, J. L. (1990). The motivational bases of public service (with L. Wise). *Public Administration Review*, 50(3), 367–373.

Perry, J. L. (2010). Revisiting the motivational bases of public service: Twenty years of research and an agenda for the future, (with A. Hondeghem, L. R. Wise). *Public Administration Review*, 70(5), 681–690.

Porter, M. E., & Kramer, M. R. (2011, January–February). Creating shared value. *Harvard Business Review*.

Part I
**The Design of Outcome-Based Performance
Management Systems in the Public Sector**

Chapter 1

Shifting from Output to Outcome Measurement in Public Administration-Arguments Revisited

Tomi Rajala, Harri Laihonen and Jarmo Vakkuri

Abstract Moving to outcome-based measurement systems in the public sector has been difficult. In this article, we examine *the contingent decision-making arguments stimulating output instead of outcome measurement in public management*. Based on an argumentative literature review, we conclude that there exist several contingent arguments encouraging politicians and public managers to stick with outputs while ignoring outcomes in performance measurement. Mapping out these arguments contributes to understanding the difficulties in implementation of outcome-based measurement and management systems. This understanding is highly useful in performance management research and policy practice. We also suggest that these contingent arguments may be considered proposals for the future research in the area of public financial management and public sector performance measurement.

Keywords Outcomes · Outcome-based performance measurement systems · Politicians · Public managers · Contingent arguments

1.1 Introduction

Outcome information is relevant to the public sector because it reports whether or not public services are producing desired outcomes to the society (Hatry 2005). This information is important to public managers seeking to improve performance as well as to other stakeholders such as voters and politicians aiming for a better societal welfare. However, it has remained extremely complicated to establish an

T. Rajala (✉) · H. Laihonen · J. Vakkuri
School of Management, University of Tampere, 33014 Tampere, Finland
e-mail: tomi.rajala@uta.fi

H. Laihonen
e-mail: harri.laihonen@uta.fi

J. Vakkuri
e-mail: jarmo.vakkuri@uta.fi

outcome-oriented measurement system for public sector purposes and practices. Governmental organizations continue to use output measures more often than outcome measures (Ferlie et al. 2005). This study aims to explore contingent arguments as to why output measures are sometimes preferred over outcome measures in the public sector.

There are many contingent arguments intrinsic to public sector behavior and performance measurement that have been acknowledged and addressed by the previous research literature (e.g., Smith 1996). However, the previous research has not been able to systematically and comprehensively understand contingent decision-making arguments for resisting the shift from an output-based to an outcome-based measurement system. Our study aims to fill this research gap by gathering these arguments together and presenting them under two topics: (1) pursue of value for money (second section) and (2) control of legitimacy (third section). In the value for money section, we are searching for arguments indicating that output information would provide more value for money than outcome information because the costs are bigger and/or benefits are not so evident in the latter. In the chapter, dealing with control of legitimacy, we are looking for arguments implicating that output information would provide more control over legitimacy than outcome information. Since legitimation (Bouckaert 1993) and value for money (Jackson 2012) are important parts of performance information use, this approach can be seen as justified.

The research follows constructivist epistemology (e.g., Guba and Lincoln 1998) and the logic of abductive reasoning (e.g., Peirce 1998). The contingent arguments are constructed from scientific arguments presented in performance management literature. As an example of our method, consider the following scenario: “scientist x has noted in her research that outcome measurement is not supported by the current entity-based information systems, and scientist y has stated that the current information systems support output measurement.” From these statements, we form a contingent argument stating that current information systems support output measurement and do not support outcome measurement. By forming this argument, we would create one possible answer to our research question. In the conclusion section, we place all these arguments under broader categories constructed in this study.

We conducted an argumentative literature review in order to construct these arguments. An argumentative literature review examines literature selectively in order to support an argument already established in the literature. The aim of this type of literature review is to develop a body of literature that establishes a contrarian viewpoint (Kennedy 2007). The contingent arguments presented in this article form a contrarian viewpoint to outcome measurement advocates listing the benefits of outcome measuring (see, e.g., Hatry 2005) compared to output measuring. These arguments describe mental models that argue against the use of outcome measures and favor output indicators. By “contingent” it is indicated that the truth value of every argument is contextual, not universal. Furthermore, these arguments are not meant to be normative in any way, and their truth value may even be untrue. The point of this article is to raise discussion on whether or not outcome measurement can have negative effects in public sector.

As a main theoretical contribution, this research gathers together the dispersed arguments describing the possible reasons why output measuring is often more established than outcome measurement in the public sector. These reasons are described in the contingent arguments, and they can be understood as problems and limitations that incentivize public sector actors not to adopt outcome measurement. Understanding of these reasons is one of the first steps in better comprehending non-use of outcome measurement. The second significant theoretical contribution is the recognition of the future research questions proposed in this study. We are hoping that future research would examine empirically whether or not these arguments are capable of explaining why the implementation of the outcome measurement has been difficult.

Figure 1.1 depicts the structure of this article. Following the introduction, first, we examine whether or not outcome measurement provides less value for money

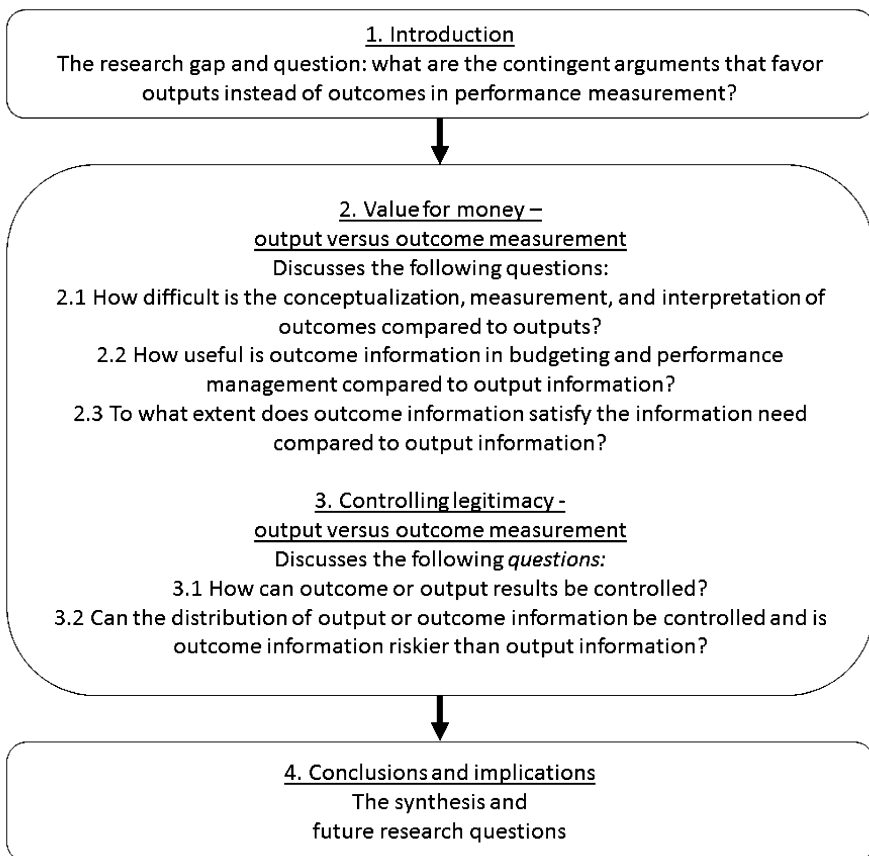


Fig. 1.1 The overall structure of the research

than output measurement. Second, we investigate how legitimacy is affected by these two types of measurement. The final part includes conclusions and future research questions.

1.2 Value for Money—Output Versus Outcome Measurement

Value for money refers to the ideal combination of whole-life costs of public services and fitness for the purpose of meeting the user's requirements (Jackson 2012). Both the outcome and output measurement have a purpose and cost. The important question in the context of this article is whether or not output measurement could provide more value for money than outcome measurement. More value for money would here indicate that output measures are cheaper to produce and/or they offer more fitness for purpose than outcome measures according to scientific arguments. Thus, in this section, we are displaying arguments presented in the literature which state that the costs of outcome measurement can be high. We are also representing previous academic statements asserting that the costs of output measuring are often low. In addition, we present ideas expressing that it might not be possible to determine the purpose of producing outcome information, or the fitness for purpose may be lacking when outcomes are attached to the performance management system and budgeting system. Thus, the use of outputs can seem more favorable in these situations to public managers and politicians.

1.2.1 Outputs Versus Outcomes: The Conceptualization, Measurement, and Interpretation

As noted by Hatry (2006), two different types of outcome exist: intermediate and end outcomes. Intermediate outcomes lead to the ends desired, but they are not ends in and of themselves. The end outcomes are the desired results of the program according to the program customers and citizens (Hatry 2006). Vedung (1997) identifies an additional outcome type: immediate outcomes. These outcomes happen right after the actions are taken, whereas intermediate outcomes occur in the causal chain following immediate outcomes (Vedung 1997). However, in Hatry's (2006) typology, immediate outcomes can be placed under the concept of intermediate outcomes without breaking any theoretical assumptions of intermediate outcomes.

Vedung (1997) has also recognized more comprehensively the complex nature of outcomes, naming several different outcome types as follows:

1. Outcomes for customers and society.
2. Quantitative and qualitative outcomes.

3. Subjective and objective outcomes.
4. Short- and long-run outcomes.
5. External outcomes and internal outcomes.
6. Positive and negative outcomes.
7. Expected and unexpected outcomes.
8. Intended and unintended outcomes.

Yeung and Matheison (1998) identify six different areas where outcomes can appear: economic performance, competitiveness, education, health, environment, democracy, and freedom. Outcomes can also occur at the program/service level, the agency/organizational level, the state/community level, or any combination thereof (e.g., Martin 1997). Thus, some outcomes for public sector bodies may occur at a societal rather than at an organizational level (McGill 2001).

The public sector has multiple outcomes at different hierarchical levels; meanwhile, different units on the same hierarchical level can consider different outcomes to be important to them. Ideally, a goal congruency exists among these different goals set by units within the public sector organization, but this might not always be the case. For example, the General Accounting Office (1997, p. 6) has reported that “mission fragmentation” is common at the federal government level in the USA, and it is difficult to get stakeholders to think beyond their own program operations to the diversity of activities related to the common outcome.

The complex nature associated with conceptualizing the outcome becomes evident when all of these outcome categories and areas are combined in the public sector’s hierarchical and horizontal dimensions. Bounded rationality and lack of know-how can magnify the complications associated with outcome definition. For example, researchers have documented problems in understanding the difference between outputs and outcomes (e.g., Dugan and Herson 2002). As Hogwood and Gunn (1992, p. 17) point out, the distinction between outcomes and outputs is often blurry in practice.

Outputs describe what the public sector does (Rosen 1993), whereas outcomes describe the effects that have been caused directly and indirectly by the outputs (e.g., Talbot 2010). The decision considering outputs of a program boils down to the following question: What are the goods and services the public sector wants to produce? No matter what the output is, all the different kinds of outcomes listed above in various areas can occur. By comparing one output to another output, only two things are compared. However, whenever the possible outcomes of the two outputs are compared, the comparison becomes far more complex. Identifying all the relevant outcomes can take more time and effort than the output identification. Value for money may not be achieved because the cost of conceptualization is too high in outcome measurement.

If outcomes cannot be defined, they cannot be measured. A key question is whether or not decision-makers agree on which types of outcomes are the most optimal and which ones can be ignored. In the public sector, a high level of subjectivity often relates to outcomes, and thus, even reaching consensus on outcomes can be difficult (e.g., Kurunmaki and Miller 2011) because outcomes can be

multi-dimensional, qualitative by nature, and impossible to represent as a single quantitative measure (Carlin and Guthrie 2003). The difficulties in defining the outcomes are well known in different countries (e.g., Carlin and Guthrie 2003). For instance, Heinrich (2002) learned that federal managers in the USA considered the outcome-based performance management systems to be conceptually and practically one of their most difficult tasks to complete. By comparison, outputs often are easier to identify (e.g., Bandy 2011, p. 76). Moreover, usually outputs have to be defined because public sector produces outputs. However, it is not required to identify and name the important and unimportant outcomes in order to get the public production up and running.

It is common that different political parties strive for different societal outcomes (e.g., Spoon and Kluver 2014). Political outcome goals may inhibit outcome measuring if these outcome goals are contradictory. The inconsistency between the policy objectives set by politicians and the goals of executive agencies also creates problems in the public sector (e.g., Smith 1995). The policy objectives are contested both among politicians as well as between politicians and managers (Agranoff and McGuire 2001). There is often little consensus as to what constitutes outcome because the large number of diverse stakeholders in the public services holds different expectations toward these services (Smith 1996). The complex nature of outcomes can intensify this rivalry whenever outcomes are policy objectives instead of outputs. The increased number of options in conceptualization simply offers more possibilities for disagreement. Investing resources in outcome measurement can lead to conflicts and inefficient resource use if outcomes cannot be defined or measured. If this scenario occurs, the purpose of outcome measurement cannot be identified properly.

Contingent argument: outputs are easier, cheaper, and less time-consuming to define and conceptualize than outcomes (nature of outcome, nature of output, and conflict orientation).

1.2.1.1 Technical Aspects of Measurement

Measuring outcomes can be astonishingly difficult (Smith 1996), and on the other hand, calculating outputs is usually fairly straightforward (Newcomer 2007). Obtaining information about the intervening variables affecting outcomes causes problems (Miller and Fox 2007), whereas variables that influence output production can be monitored and detected more easily in many cases. The problem with intervening variables in the context of outcomes is related to the problem of monitoring citizens and societal activities round-the-clock holistically. Anthony and Young (1988, p. 608) summarize this common problem that plagues many outcome performance measures (here “social indicator” means “outcome”):

A social indicator is a broad measure of output which significantly reflects the work of the organization. Unfortunately, few social indicators can be related to the work of a single organization because in almost all cases they are affected by exogenous forces; that is,

forces other than those of the organization being measured. The crime rate in a city may reflect the activities of the police department and the court system, but it is also affected by unemployment, housing conditions, and other factors unrelated to the effectiveness of these organizations... Social indicators are so nebulous, so difficult to obtain on a current basis, so little affected by current program efforts, and so much affected by external influences that they are of limited usefulness in day-to-day management...

In public and private contexts, some outcomes cannot be measured directly, and some outcomes are not measurable at all (e.g., van der Valk and van Iwaarden 2011; Newcomer 2015). In such cases, output measurement has to suffice (Cunningham and Harris 2001). In addition, factors related to the reliability, validity, and accuracy of the measurement may favor output instead of outcome measurement. As Mcphee (2005) points out, the reported information on output tends to be better than for outcomes because output indicators are often more appropriate and the method for output measurement is usually more robust and reliable than for outcomes. Outcomes are often encompassed by values of quality and satisfaction (e.g., Chalmers 2008). They are considered to be more difficult to measure than outputs (Currstine et al. 2008), which are often more quantitative by nature. As a consequence, outcomes are not utilized nearly as often as outputs in practice (e.g., OECD 2013).

One permanent problem with outcomes is that the impact of any governmental action requires information about what would have happened to citizens if those actions were not executed (e.g., Heinrich 2002). When assessing the effectiveness of government actions, it is difficult to isolate and measure the real difference between doing something and doing nothing. Again, unmeasured intervening variables and moderator variables can explain outcomes better than measured ones. On the contrary, doing something and doing nothing can be seen rather easily on production volumes (e.g., Rosen 1993).

Outputs are usually cheaper to measure (e.g., Marks 2005). In contrast, measuring all the relevant aspects of the outcomes would normally require rigorous quantitative and qualitative methodology with subjects over prolonged time periods (Schalock 2001). In such cases, outcome measurement can require extensive resources or tunnel vision focusing on some aspects while ignoring other critical aspects associated with outcomes by reducing the number of indicators used to track outcome development in the name of measurement efficiency (Lowe 2013). If extensive outcome measurement is chosen, frontline workers will often have to devote more time to reporting and less time to service production (e.g., Keevers et al. 2012). The question here might simply be whether we want to focus on reporting or on the actual service production.

Conflicts about the usefulness of different approaches to public sector performance measurement do exist (Harrison et al. 2012). Agreeing on appropriate performance measures has proven difficult in hybrid organizations (Kurunmaki et al. 2003). Performance measures are generally not neutral in the public sector context (e.g., Van de Walle and Van Dooren 2010), and there exist divergent opinions about the right performance indicators among politicians and between politicians and managers (Agranoff and McGuire 2001). The development of political debate

dictates the assessment of public sector performance (Stewart and Walsh 1994). Because there can be a lack of consensus regarding the right indicators, performance measurement can cause dysfunctional effects (van Thiel and Leeuw 2002).

According to Chan (2004), outcome measures often are more difficult to define than output measures. Lack of consensus from the right outcome measures often occurs (Newcomer 2015). Again, the complexity of outcomes offers more possibilities to measure, meaning that there are more alternatives from which to choose the performance indicators. The diversity of preferences typical to public sector can utilize these alternatives to create conflicts. These conflicts can induce more costs and mean that the purpose of outcome measurement cannot be defined.

Contingent argument: outcome measurement causes more costs and conflicts about the right measures than output measurement. Meanwhile, outcomes cannot be measured comprehensively, whereas outputs can be (nature of outcome, nature of output, and conflict orientation).

1.2.1.2 Interpretation Problems in Outcome Results

The analysis of causes explaining the outcomes is often more complicated than the analysis of the activities producing the outputs (e.g., Pollitt and Bouckaert 2004). The linkage and interaction between outcomes, outputs, intervening variables, and/or moderator variables makes the interpretation of outcome results considerably more difficult (e.g., Mascarenhas 1996). How different policies, programs, and agencies contribute to outcomes is often unclear (Newcomer 2015). The fact that a perception in a complex issue depends on when, where, and who is making the interpretation does not help in outcomes analysis (e.g., Kunda 1990; Van Maanen and Schein 1979). These complications in detecting the causes explaining the outcomes are called “the attribution problem” in the previous literature, and several researches have addressed this problem (e.g., Taro 2015).

Complex outcomes may cause information overload for politicians and public managers and therefore deteriorate the quality of decisions (c.f. Hahn et al. 1992). For this reason, simpler output information may seem a better choice (e.g., Chaston 2011). Kristensen et al. (2002) point out that politicians and public managers can devote focused attention to only limited areas at a time, and these actors have constraints on how much information they can utilize in their decision-making. If the outcomes form from complex processes, the decision-maker may not be able to utilize all of the information relating to the outcome achievement. Outcome measurement may not provide sufficient value for money if it deteriorates the quality of decisions or the information remains unused because we cannot interpret it properly or without conflicts and debates.

Contingent argument: interpretation of outcomes is more difficult and more prone to produce conflicts than interpretation of outputs (nature of outcome, nature of outputs, and conflict orientation).

1.2.1.3 Output as a Reflection of an Outcome

When outputs reflect outcomes reasonably well, the value of producing additional outcome measures can be very low. It is therefore important to analyze how well current outputs can approximate outcomes of a public organization (e.g., Smith 1996). For example, in the private sector, there often is no need to measure customer outcomes because the customers' valuation of the products and services reflects their willingness to pay for them (Smith 1996). In a similar fashion, the willingness to use, for example, the public sport facilities or the public parks can tell us something about the customer valuation placed on these types of goods and services.

Contingent argument: outputs reflect outcomes adequately and accordingly; there is no need for outcome information (nature of outcome and nature of output).

1.2.2 How Outputs and Outcomes Connect to Budgeting and Performance Management

The budget demonstrates whether or not there is political and managerial demand for outcome measurement (c.f. Greenwood et al. 1977). The budget process reminds us that there are opportunity costs for measuring outcomes. The interesting question in the resource allocation context is what makes output measurement more desirable than outcome measurement. The answer is threefold, relating to costs, current information systems, and the purpose of such systems.

From the perspective of budgetary allocations, the decision-making problem is about comparing uses of resources to the added value of measurements. Assuming that the added value is perceived to be similar between the two types of measurement, cost of measurement defines the choice. Performance information often focuses on output levels because these are easy and less costly to define, measure, and analyze. By comparison, program outcomes tend to be much more difficult to identify, measure (e.g., Robichau and Lynn 2009), and analyze (e.g., Mascarenhas 1996). For these reasons, outputs may be preferred. If the outcome information is more expensive than output information, the former would have to provide more value than the latter in order to be the first choice of the decision-maker when these two types of measurement are competing on the same resources.

Contingent argument: From the budgetary perspective output measurement may provide more value for money because the nature of outcomes is problematic and more expensive to measure and analyze comprehensively (opportunity costs/competition for resources).

The necessity of coupling the budget process to outcomes may be one reason explaining why output measuring is preferred over outcomes in the public sector. As Kristensen et al. (2002) stated government budgeting and financial systems may currently only be capable of generating rudimentary matches of resources and

outcomes. It is difficult to put a price tag on outcomes (e.g., Midwinter 2009). In contrast, calculating the cost of achieving required output levels is a rather established procedure (e.g., Anderson et al. 2000). There is thus a twofold problem with connecting outcomes to budgets. The problem with cost calculation is more fundamental because it has to be solved before any information system can be built. However, the problem involving information systems should not be understated either. Performance measurement in the public context is often based around the traditional vertical hierarchies of government departments and developed within individual legal entities (e.g., Ryan and Walsh 2004). Outputs are often created in these entities, whereas end outcomes are not (e.g., Mayne 2007). Systems supplying information on the costs and benefits of working across accounting entities would be needed in order to do pooled budgets that assign resources to service outcomes, for instance (Hodges 2012). Thus, information systems may not support outcome measurement.

Contingent argument: outcome measurement requires too many investments in information systems, while output measurement does not (information system).

Difficulties in cost calculations mostly relate to the fact that outcomes may not be as accurately specified and measured as outputs. Also, the causal link between inputs and outcomes is often more difficult to perceive than the link between inputs and outputs. Thus, uncertainty may arise over how changes in resource levels may affect overall performance (Kristensen et al. 2002). This problem causes difficulties, especially at the state and community level. Connecting resources and outcomes to the change in indicator values in the state and community level is difficult at best, and it raises validity issues that are not encountered at the other levels (Rossi 1997). As stated by Kristensen et al. (2002), outcome budgeting raises many difficult questions. For instance, who should estimate the resources needed for outcomes that are a result of cross-sectional government operations? And should outcome targets be set first and then resources after the targets or vice versa (Kristensen et al. 2002)? These quite practical questions demonstrate the challenging problems outcome budgeting can generate (e.g., Grizzle and Pettijohn 2002).

Timeliness is an important feature of the performance management system (Heinrich 2002). The ability to provide timely feedback to public managers creates opportunities for performance improvements and for adjustments in budget allocations, service contracts, management practices, and training strategies. The challenge here is to provide outcome information in a timely manner so that it can be connected to day-to-day performance management. If outcome information cannot be used in operational performance management, the purpose of providing such information becomes compromised.

According to Heinrich (2002), federal agencies in the USA have found it particularly difficult to transform their long-term missions or strategic goals into annual performance goals. These federal agencies have also found predicting the level of performance results attained over a shorter term to be particularly challenging. For this reason, short-run rather than long-run measures are normally used in performance standards systems (Heinrich 2002). The indication here seems to be that

outcomes cannot be utilized if they require long-run measures, as they often do. A longer time frame is usually needed because evaluating how the programs have affected the outcome takes time (Bovaird 2014). Outputs, on the other hand, are more suitable for those performance management systems that aim to provide feedback with minimal lag from actual performance because outputs can usually be detected, measured, and reported more instantly and easily than outcome information. Outputs also can be used to control work more efficiently because, by defining outputs, the public managers and politicians actually define what is done at the operating level (e.g., Snell 1992). Outcomes, on the other hand, may provide more freedom to the frontline workers. For instance, it does not matter what precise actions are taken as long as customers are satisfied.

Contingent argument: it is difficult to do budgets for outcomes and use outcome information in day-to-day performance management whereas outputs can be more easily connected to budgeting and performance management processes (nature of outcomes and nature of outputs).

1.2.3 The Information Need

According to Dervin's (1983) sense-making approach, information needs arise from the gap that exists between the current situation and the desired situation, from the process that tries to make sense of the current gap and from the efforts to bridge that gap. Put simply, information needs are conceived as individual attempts to answer questions and to make sense of a gap in order to move from the current situation to the desired situation (Dervin 1983).

If low information need causes problems for the adoption of outcome measurement, the politician or public manager fails to see outcome information as beneficial for four reasons. Firstly, outcome information perhaps cannot help the politicians and public managers to understand which outputs will produce certain outcomes. In this situation, the outcome information does not provide enough data on how to change the current system; therefore, this information does not lead to action. The lack of mutual congruence in the results analysis may also mean that the information would remain unused. Secondly, it might be that the political system and public managers cannot agree on what the desired situation, or outcome, should be. Thus, there exist multiple views on the situation, which leads to an inability to determine the kind of information is needed collectively in order to improve the quality of life in society. Thirdly, the lack of know-how in performance measurement can lower information needs if it is acknowledged that these limitations could deteriorate the quality of the information to a level where it is no longer useful. Finally, the information can be seen as a blame attractor and as uncontrolled risk if outcomes are not in control and transparent information is needed for the sake of legitimacy.

Contingent argument: outcome information is costly to produce and the value to the decision-maker is equal or less than output information, which is why output information is preferred (information need).

1.3 Controlling Legitimacy—Output Versus Outcome Measurement

It has been stated that the main motivation for the use of performance information is legitimacy-seeking rather than efficiency maximization (Modell 2001). For example, in symbolic use, the information can be used for legitimation purposes (Van de Walle and Van Dooren 2008). Thus, legitimacy can be considered an important aspect of the implementation of performance measurement, and therefore, it is justified to look how produced output and outcome information can be managed and controlled in order to gain legitimacy. Suchman (1995, p. 574) describes *legitimacy* as “a generalized perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions.” To Zimmerman and Zeitz (2002), this social system is a very broad concept that includes the operating environment in which the organization functions and which it needs to demonstrate consistency. Legitimacy can also mean the congruence between the organization’s activities and outcomes and society’s values, norms, and expectations (Ashforth and Gibbs 1990). In order to understand how legitimacy can be managed, it is important to take a more detailed look at the ability to control the output and outcome achievement, as well as the distribution of the performance information describing these achievements.

1.3.1 Controllability of Results

Gaining, maintaining, and repairing legitimacy may be more problematic when outcomes are reported instead of outputs. This difficulty rests in the fact that outputs often have higher controllability than outcomes (Irwin 1996).

Contingent argument: due to higher controllability earning legitimacy with output rather than outcome information renders more control (control of legitimacy).

Lack of control also leads to problems in accountability. It seems that using outputs in performance management could lead to situations where the government has fewer difficulties holding an agency accountable for delivering the agreed-upon outputs (Mayne 2007). Outcome measures, on the other hand, may be subject to multiple determinants, with the budget holder’s activities representing just one. It is harder for the government to hold the agency accountable for outcome achievement if an agency has only partial control over outcomes (Mayne 2001). In a similar fashion, ministers cannot be held accountable if the outcomes are not within their control (e.g., Irwin 1996). In addition, the time frame related to outcomes is

troublesome because there can be extensive time lags between resource use and performance outcomes (e.g., Bovaird 2014). Consequently, the attribution problems not only make it difficult to interpret outcome results but also produce problems to accountability (Mayne 2001).

Attribution of responsibility for outcomes becomes even more problematic when the services are supported by multiple funding sources or various providers, such as health service providers, measures are affected by so many determinants that change in outcomes cannot be attributed to the effectiveness of a specific program alone (U.S. Department of Health and Human Services 1997). The uncontrolled nature of outcomes can also raise another challenging question: whether or not it is politically or professionally wise to promise to deliver outcomes instead of outputs. It can be daunting to manage efforts to achieve outcomes that require actions across different agencies or will be achieved by more than one hierarchical level of government (Kristensen et al. 2002).

Contingent argument: while outputs are under control, politicians or public managers have only partial control over outcomes and only partial accountability for the results. Thus, outcomes may offer less tools for principals to control performance (control of accountability).

1.3.2 Controllability of Information

Van de Walle and Van Dooren (2010) note that information relates to power structures because any new information about the performance of organizational departments may have a significant effect on future budgets or staff allocations. Even the survival of the department within the wider organization can depend on performance information. It is therefore in organizational actors' interests to control information flows (Van de Walle and Van Dooren 2010).

If government agencies were to focus on outcomes, the stakeholders should understand that the agency is only one of many factors likely to affect outcomes (e.g., Schalock 2001). This recognition would explicitly indicate that public agencies have only partial control and, therefore, only partial accountability, according to the stakeholders (Hatry 1997, p. 2). Without this recognition from stakeholders, blame games and blame avoidance strategies will most likely play a role in performance management. However, the opposition versus government setting can prevent such recognition because political opposition can do little other than generate blame. They cannot hope to have an effective voice in the process of policy formulation so long as there is a majority government in the parliamentary system (e.g., Weaver 1986). Thus, whenever there exists opposition, there may also exist a need to control information.

It is an interesting question whether or not there is a larger need to control outcome information than information that describes outputs. To answer this question, we need to ask: What do the outcome and output information tell us, exactly? If the output goals

are not met, it indicates that the government is not operating efficiently. However, it cannot be inferred from output levels that the government is doing the wrong things. Failing to achieve outcome goals more directly raises the question of whether or not the government is actually doing the wrong things. Thus, the comparison between output and outcome measurements relates to the comparison between efficiency and effectiveness, where efficiency is “doing things right” and effectiveness is “doing the right thing” (e.g., Gleason and Barnum 1982).

Doing the wrong thing is a more severe error than doing the right thing inefficiently. Taking the wrong actions not only wastes public resources but can also lower citizens’ well-being. Thus, outcome indicators have the potential to show more fundamental problems in government operations than output indicators. Moreover, doing the wrong things demonstrates problems in the political system and in the current government’s visions. For politicians and public managers, the rationale may be to think that nothing is worse than providing outcome information demonstrating that public sector is doing the wrong things. However, it is uneasy to demonstrate that the government is actually doing the wrong things because of the inherent ambiguity analysis of outcomes.

From the perspective of accountability, outputs provide no justification for failures. This observation could indicate that the agent accountable for the outputs would have a greater stake because no excuses for failure would be available when output levels are not achieved. If the existence of blame avoidance is assumed, there is a need to control output information. Outcomes, on the other hand, offer only partial control over results; however, this partial control also offers justification to fail. Understanding the nature of outcomes would therefore diffuse the blame, avoiding the need to use blame avoidance strategies. Thus, it is unclear which type of information can be a bigger threat to legitimacy. Ultimately, if people react to reality as they perceive it and not to reality itself (Lewin 1936), then the need to control different information types depends upon stakeholders’ reactions. In general, citizens tend to attach outputs and outcomes to specific programs (Taro 2015). If so, then we are back to comparing the harmful consequences of effectiveness and efficiency information.

According to Wholey and Hatry (1992), public managers fear that elected officials, interest groups, and the media may use outcome information as fodder for attacks. The possible misuse of negative findings is a risk that comes with performance information (Wholey and Hatry 1992). This fear is not unjustified because it is a common phenomenon in politics (and in human behavior) that negative information produces more activity and impact than positive information (Rozin and Royzman 2001). This negativity bias encourages the avoidance of bad publicity and can influence the willingness to provide performance information.

Because outcomes are not under the control of politicians and public managers, transparency can generate bad publicity and adverse effects by putting poor results in the spotlight every time a partially uncontrollable outcome goal is not achieved. The effects of poor results depend on whether the public sector is applying full transparency (c.f. Rousseau 1772), direct transparency (c.f. Bryan 2010), or indirect transparency (c.f. Hood 2007). For these reasons, the ability to control information is linked closely to the chosen state of transparency.

Information can be controlled in two ways: by inhibiting information production or by controlling what information is delivered and to whom. The inability to determine who gets the information may lead to situations where outcome measuring will be inhibited because this is the only way to ensure that information about the negative results does not end up in the hands of opponents.

*Contingent argument: outcome information may expose more severe errors in public sector actions than output information. The inability to control the distribution and production of the outcome information in a transparent setting may attract too much blame and trigger dysfunctional behavior, conflicts, and blame games (**controllability of information and conflict orientation**).*

By evaluating the credit-claiming and blame-avoiding opportunities in different situations, it becomes apparent that politicians or public managers may choose to be loss averse, risk averse, risk neutral, or risk seeking. Depending on this choice, the arguments introduced in this research can be valued differently. For example, a risk-seeking politician might not care about the possibility of outcome measurement producing bad publicity or conflicts in the institution; by comparison, loss-averse politicians may care a great deal and make a choice accordingly. Typically, people are more loss averse (Tversky and Kahneman 1992), and politicians often choose to be risk averse (Weaver 1986). The constituencies and beneficiaries may also prefer that the results of the government program go unmeasured because this measurement could demonstrate that the program has actually been ineffective, of little value, or unimportant in achieving the desired effect or impact in the society (Kristensen et al. 2002).

*Contingent argument: loss-averse politicians and public managers try to avoid conflicts, professional and political disasters, resource wasting, and legitimacy losses. They will not promise to deliver outcome information because outcome results may be ticking time bombs that are beyond their control, at least partially (**loss aversion**).*

1.4 Conclusions and Implications

As a main contribution, we found several contingent arguments relevant to politicians and public managers. These arguments can be examined when the transition from output to outcome measurement is undermined (see Fig. 1.2). The arguments are linked to each other, and together, they form a complex network of issues that may influence the decision-makers to reject or resist outcome-based performance management.

We organized the wide array of arguments under two more general topics. We looked at whether or not outcome measurement provides value for money and improves legitimacy. However, we are fully aware that the arguments could have been organized differently. In order to develop our theoretical–conceptual framework, we now reorganize all the contingent arguments presented in Fig. 1.2 under broader categories:

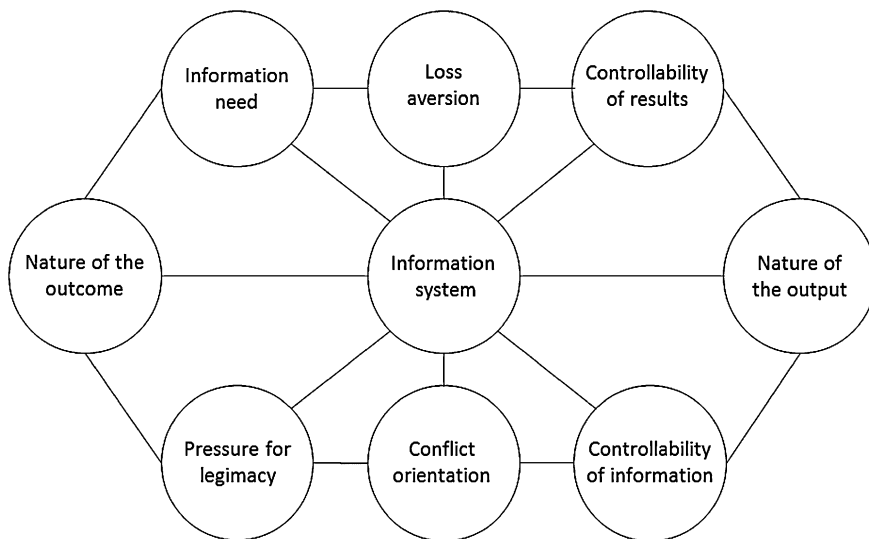


Fig. 1.2 The contingent arguments to ignore outcome measurement

1. Information (information need and information system).
2. Controllability (controllability of results and information).
3. Legitimacy.
4. Nature of outcome measurement (nature of outcomes).
5. Political conflict (competition for resources, conflict orientation, and loss aversion).

These categories can be used to assess relevant contingent arguments that may inhibit the development of outcome measurement. Past and future research on outcomes can also be classified according to the above categories.

We do not assume that all the contingent arguments are present or assessed at the point of decision-making. In fact, it is probably more likely that some arguments are not even recognized by the decision-maker. We only assume that if at least one contingent argument is acknowledged and considered by politicians or public managers, outcome measurement may be rejected. We also acknowledge that these arguments can be valued differently by various politicians and public managers, and this valuing most likely varies among different decision-making situations. We also do not exclude the possibility of emergence: The whole could somehow differ from the parts (c.f. Morowitz 2002), and the final decision may deviate from the decision made purely based on weighting and calculating all the arguments, either favoring or opposing outcome measurement.

Finally, these contingent arguments can potentially enrich our theoretical and analytical view of the institutional practices and problems of developing outcome

Table 1.1 The future research questions

Future research questions
What type of information need can the outcome information satisfy, and how should the use and context of use be designed based on this need? (information need)
How would the current information systems have to be expanded or transformed if cross-sectional outcome measurement is to be connected to day-to-day management? (information system)
Do voters, political opposition, and other stakeholders understand that the government has only partial control over the outcomes, or do they exploit bad outcome results in order to gain advantage in elections and political decision-making by using blame games? (controllability of results)
Does transparency positively or negatively affect incentives to measure outcomes? (controllability of information)
Are voters, politicians, and public managers attaching outcome measurement to legitimacy in general? (control of legitimacy)
Do politicians, public managers, and voters recognize the complexity of outcomes and the political power associated with outcome measurement? (nature of outcome)
In which situations can outputs reflect outcomes adequately and accordingly? (nature of outputs)
How high is outcome measurement in the hierarchy of needs when resource allocations are considered by politicians and public managers? (opportunity costs/competition for resources)
How does a conflict-oriented environment affect incentives to measure outcomes? (conflict orientation)
Are public managers and politicians risk averse, loss averse, risk neutral, or risk-loving when it comes to deciding whether or not to implement the outcome measurement? (loss aversion)

measurement in public administration. Therefore, we suggest that these contingent arguments be taken as proposals for the future research endeavors in the area of public financial management and public sector performance measurement (see Table 1.1). If these contingent arguments are supported by the empirical evidence in future research, they can be obstacles preventing the implementation outcome measurement. Taken into consideration the importance of outcome information to the stakeholders of the public sector, this threat cannot be taken lightly.

References

Agranoff, R., & McGuire, M. (2001). American federalism and the search for models of management. *Public Administration Review*, 61(6), 671–681.

Anderson, G. F., Hurst, J., Hussey, P. S., & Jee-Hughes, M. (2000). Health spending and outcomes: Trends in OECD countries, 1960–1998. *Health Affairs*, 19(3), 150.

- Anthony, R. N., & Young, D. W. (1988). *Management control in nonprofit organizations*. Homewood, IL: Irwin.
- Ashforth, B. E., & Gibbs, B. W. (1990). The double-edge of organizational legitimation. *Organization Science*, 1(2), 177–194.
- Bandy, G. (2011). *Financial management and accounting in the public sector*. Abingdon, OX: Routledge.
- Bouckaert, G. (1993). Measurement and meaningful management. *Public Productivity and Management Review*, 17(1), 31–43.
- Bovaird, T. (2014). Attributing outcomes to social policy interventions—‘Gold standard’ or ‘fool’s gold’ in public policy and management? *Social Policy and Administration*, 48(1), 1–23.
- Bryan, F. M. (2010). *Real democracy: The New England town meeting and how it works*. Chicago: University of Chicago Press.
- Carlin, T., & Guthrie, J. (2003). Accrual output based budgeting systems in Australia The rhetoric-reality gap. *Public Management Review*, 5(2), 145–162.
- Chalmers, D. (2008). *Indicators of university teaching and learning quality*. Canberra: Australian Learning and Teaching Council, the Australian Government Department of Education.
- Chan, Y. C. (2004). Performance measurement and adoption of balanced scorecards: A survey of municipal governments in the USA and Canada. *International Journal of Public Sector Management*, 17(3), 204–221.
- Chaston, I. (2011). *Public sector management: Mission impossible?* Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.
- Cunningham, G. M., & Harris, J. E. (2001). A heuristic framework for accountability of governmental subunits. *Public Management Review*, 3(2), 145–165.
- Curristine, T., Park, C. K., & Emery, R. (2008). Budgeting in Portugal. *OECD Journal on Budgeting*, 8(3), 7.
- Dervin, B. (1983). Information as a user construct: The relevance of perceived information needs to synthesis and interpretation. In S. Ward & L. Reed (Eds.), *Knowledge structure and use: Implications for synthesis and interpretation* (pp. 153–184). Philadelphia: Temple University Press.
- Dugan, R. E., & Heron, P. (2002). Outcomes assessment: Not synonymous with inputs and outputs. *The Journal of Academic Librarianship*, 28(6), 376–380.
- Ferlie, E., Pollitt, C., & Lynn, L. E. (2005). *The Oxford handbook of public management*. Oxford: Oxford University Press.
- General Accounting Office. (1997). *Managing for results: Opportunities for continued improvements in agencies performance plans*. Washington, DC: Government Printing Office. GAO/AIMD-99-215.
- Gleason, J. M., & Barnum, D. T. (1982). Toward valid measures of public sector productivity: Performance measures in urban transit. *Management Science*, 28(4), 379–386.
- Greenwood, R., Hinings, C. R., & Ranson, S. (1977). The politics of the budgetary process in English local government. *Political Studies*, 25(1), 25–47.
- Grizzle, G. A., & Pettijohn, C. D. (2002). Implementing performance based programme budgeting. *A System-Dynamics Perspectives, Public Administration Review*, 62(1), 51–62.
- Guba, E. G., & Lincoln, Y. S. (1998). Competing paradigms in qualitative research. In N.K. Denzin & Y.S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd edn., pp. 191–215). Thousand oaks: Sage Publications.
- Hahn, M., Lawson, R., & Lee, Y. G. (1992). The effects of time pressure and information load on decision quality. *Psychology and Marketing*, 9, 365–378.
- Harrison, J., Rouse, P., & De Villiers, C. J. (2012). Accountability and performance measurement: A stakeholder perspective. *Journal of Centrum Cathedra: The Business and Economics Research Journal*, 5(2), 243–258.
- Hatry, H. P. (1997). Where the rubber meets the road: Performance measurement for state and local public agencies. *New Directions for Evaluation*, 1997(75), 31–44.
- Hatry, H. P. (2006). *Performance measurement: Getting results*. The Urban Institute: Washington, DC.

- Hatry, H. P. (2005). Results matter: Suggestions for developing country's early outcome measurement effort. In A. Shah (Ed.), *Handbook on public sector performance review*. Washington, DC: World Bank.
- Heinrich, C. J. (2002). Outcomes-based performance management in the public sector: Implications for government accountability and effectiveness. *Public Administration Review*, 62(6), 712–725.
- Hodges, R. (2012). Joined-up government and the challenges to accounting and accountability researchers. *Financial Accountability and Management*, 28(1), 26–51.
- Hogwood, B. W., & Gunn, L. A. (1992). *Policy analysis for the real world*. Oxford: Oxford University Press.
- Hood, C. (2007). What happens when transparency meets blame-avoidance? *Public Management Review*, 9(2), 191–210.
- Irwin, T. (1996). *An analysis of New Zealand's new system of public sector management. Public management occasional papers no 9. Performance management in government. Contemporary Illustrations*. Paris: Organisation for Economic Co-operation and Development.
- Jackson, P. (2012). *Value for money and international development: Deconstructing myths to promote a more constructive discussion*. Paris: Organisation for Economic Cooperation and Development.
- Kennedy, M. M. (2007). Defining a literature. *Educational Researcher*, 36(3), 139–147.
- Kunda, Z. (1990). The case for motivating reasoning. *Psychological Bulletin*, 108, 480–498.
- Keevers, L., Treleaven, L., Sykes, C., & Darcy, M. (2012). Made to measure: Taming practices with results-based accountability. *Organization Studies*, 33(1), 97–120.
- Kristensen, J. K., Groszyk, W., & Bühler, B. (2002). Outcome-focused management and budgeting. *OECD Journal on Budgeting*, 1(4), 7–34.
- Kurunmaki, L., Miller, P., Keen, J., & Institute of Chartered Accountants in England and Wales. (2003). *Health act flexibilities: Partnerships in health and social care*. London: Centre for Business Performance.
- Kurunmaki, L., & Miller, P. (2011). Regulatory hybrids: Partnerships, budgeting and modernising government. *Management Accounting Research*, 22(4), 220–241.
- Lewin, Kurt. (1936). *Principles of topological psychology*. New York: McGraw-Hill.
- Lowe, T. (2013). New development: The paradox of outcomes—the more we measure, the less we understand. *Public Money and Management*, 33(3), 213–216.
- Marks, M. M. (2005). In S. Mathison (Ed.), *Encyclopedia of evaluation (287)*. Thousand Oaks, CA: Sage.
- Martin, L. L. (1997). Outcome budgeting: A new entrepreneurial approach to budgeting. *Journal of Public Budgeting, Accounting and Financial Management*, 9(1), 108.
- Mascarenhas, R. C. (1996). Searching for efficiency in the public sector: Interim evaluation of performance budgeting in New Zealand. *Public Budgeting and Finance*, 16, 13–27.
- Mayne, J. (2001). Addressing attribution through contribution analysis: Using performance measures sensibly. *The Canadian Journal of Program Evaluation*, 16(1), 1.
- Mayne, J. (2007). Challenges and lessons in implementing results-based management. *Evaluation*, 13(1), 87–109.
- McGill, R. (2001). Performance budgeting. *International Journal of Public Sector Management*, 14(5), 376–390.
- McPhee, I. (2005). Outcomes and outputs: Are we managing better as a result?. In *CPA National Public Sector Convention*, 20.
- Miller, H. T., & Fox, C. J. (2007). *Postmodern public administration*. New York: M.E. Sharpe.
- Midwinter, A. (2009). New development: Scotland's Concordat—An assessment of the new financial framework in central–local relations. *Public Money and Management*, 29(1), 65–70.
- Modell, S. (2001). Performance measurement and institutional processes: A study of managerial responses to public sector reform. *Management Accounting Research*, 12(4), 437–464.
- Morowitz, H. J. (2002). *The emergence of everything: How the world became complex*. Oxford: Oxford University Press.

- Newcomer, K. E. (2007). Measuring government performance. *International Journal of Public Administration*, 30(3), 307–329.
- Newcomer, K. E. (2015). From outputs to outcomes. In Guy & Rubin (Eds.) *Public administration evolving: From foundations to the future* (pp. 125–158). New York: Routledge.
- OECD. (2013). *Together for better outcomes: Engaging and involving SME taxpayers and stakeholders*. Paris: Author.
- Peirce, C. S. (1998). *The essential Peirce selected philosophical writings*. Bloomington, IN: Indiana University Press.
- Pollitt, C., & Bouckaert, G. (2004). *Public management reform: A comparative analysis*. Oxford: Oxford University Press.
- Robichau, R. W., & Lynn, L. E., Jr. (2009). The implementation of public policy: Still the missing link. *Policy Studies Journal*, 37(1), 21–36.
- Rosen, E. D. (1993). *Improving public sector productivity: Concepts and practice*. Newbury Park, CA: Sage Publications.
- Rossi, P. (1997). Program Outcomes: Conceptual and measurement issues. In E. Mullen & J. Magnabosco (Eds.), *Outcomes measurement in human services* (pp. 20–34). Washington, DC: NASW Press.
- Rousseau, J. J. (1772/1985). *The government of Poland*, translated, with an introduction and notes by Willmoore Kendall, Indianapolis: Hackett Publishing Co.
- Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, 5(4), 296–320.
- Ryan, C., & Walsh, P. (2004). Collaboration of public sector agencies: Reporting and accountability challenges. *International Journal of Public Sector Management*, 17(7), 621–631.
- Schallock, R. L. (2001). *Outcome-based evaluation*. New York: Kluwer Academic/Plenum Publishers.
- Smith, P. (1995). On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration*, 18(2–3), 277–310.
- Smith, P. C. (1996). *Measuring outcome in the public sector*. New York: Taylor and Francis.
- Snell, S. A. (1992). Control theory in strategic human resource management: The mediating effect of administrative information. *Academy of Management Journal*, 35(2), 292–327.
- Spoon, J. J., & Klüver, H. (2014). Do parties respond? How electoral context influences party responsiveness. *Electoral Studies*, 35, 48–60.
- Stewart, J., & Walsh, K. (1994). Performance measurement: When performance can never be finally defined. *Public Money and Management*, 14(2), 45–49.
- Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review*, 20(3), 571–610.
- Talbot, C. (2010). *Theories of Performance: Organizational and service improvement in the public domain*. Oxford: Oxford University Press.
- Taro, Kulli. (2015). *The attribution problem in performance measurement in the public sector: Lessons from performance audits in Estonia*. Tallinna: TUT Press.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), 297–323.
- U.S. Department of Health and Human Services. (1997). *Strategic Plan*. Washington, DC: Author.
- Van Dooren, W., & de Walle, Van. (2008). *Performance information in the public sector: How it is used*. Basingstoke, Hampshire: Palgrave Macmillan.
- Van de Walle, S., & Van Dooren, W. (2010). How is information used to improve performance in the public sector? Exploring the dynamics of performance information. In K. Walshe, G. Harvey, & P. Jas (Eds.), *Connecting knowledge and performance in public services: From knowing to doing* (pp. 33–55). Cambridge: Cambridge University Press.
- Van der Valk, W., & Van Iwaarden, J. (2011). Monitoring in service triads consisting of buyers, subcontractors and end customers. *Journal of Purchasing and Supply Management*, 17(3), 198–206.

- Van Maanen, J., & Schein, E. H. (1979). *Toward a theory of organizational socialization*. Cambridge, Mass: M.I.T.
- Van Thiel, S., & Leeuw, F. L. (2002). The performance paradox in the public sector. *Public Performance and Management Review*, 25(3), 267–281.
- Vedung, E. (1997). *Public policy and program evaluation*. New Jersey: Transaction Publishers.
- Weaver, R. K. (1986). The politics of blame avoidance. *Journal of Public Policy*, 6(04), 371–398.
- Wholey, J. S., & Hatry, H. P. (1992). The case for performance monitoring. *Public Administration Review*, 52, 604–610.
- Yeung, O. M., & Mathieson, J. A. (1998). *Global benchmarks: Comprehensive measures of development*. Washington, DC: Brookings Institution Press.
- Zimmerman, M. A., & Zeitz, G. J. (2002). Beyond survival: Achieving new venture growth by building legitimacy. *Academy of Management Review*, 27(3), 414–431.

Chapter 2

Organizational Performance in the Italian Health care Sector

Alessandro Spano and Anna Aroni

Abstract The public sector performance management (PM) literature is particularly rich as this topic is one of the most appealing for public sector scholars (Pollitt, *J Public Adm Res Theory* 6:25–44, 2005). However, organizational performance (OP) has been neglected across the world (Andrews et al. *J Public Adm Res Theory* 21:i301–i319, 2011) as well as in the Italian public administration (Martin and Spano, *Public Money Manag* 35:303–310, 2015). This chapter investigates how OP is defined, measured, and evaluated in the Italian health care sector. Our analysis showed the limited use of performance management in Italian public health organizations and a high variability in the way OP is defined and measured. This makes it difficult to compare the results of different organizations. For this reason, future standardization could allow policy makers to improve the accountability.

Keywords Organizational performance · Health care · Italian public sector

2.1 Introduction

The issue of OP is of particular relevance in the healthcare sector, where the impact of health organizations on individuals' lives is significant and measures of OP are required to understand the extent to which these organizations are effective. Even though significant progress has been made in building more advanced performance measurement systems in the health care sector, more work is needed (Smith et al. 2008). In fact, the literature on performance management in the health care sector reports several cases of incorrect uses and, even misuses of performance measures and targets with the introduction of a kind of “governance by targets” and a

A. Spano (✉) · A. Aroni
Department of Economic and Business Sciences, University of Cagliari, Cagliari, Italy
e-mail: spano@unica.it

A. Aroni
e-mail: anna.aroni@unica.it

consequent increased risk of gaming (Bevan and Hood 2006). Micheli and Neely (2010) also report a lack of coherence among the different actors involved in the setting of objectives and targets at different levels, from central to local, making performance measurement more complex.

Traditionally, performance in health care has been measured using specific indicators such as incidence of pathology, mortality measures, and measures of mortality after a specific treatment. Other measures are increasingly attractive, including those that focus on patient health status, which are often in the form of outcome measures (Smith et al. 2008). However, there is a limited “understanding of how performance measurement can be organized to support improvement initiatives in health care practices” (Elg et al. 2013).

In the Italian public sector, the role of OP has been largely neglected, and more importance has been given to individual performance (Martin and Spano 2015). As far as the healthcare sector is concerned, OP is attracting increased attention in Italy, but there is still a lack of extensive research on this topic. For example, there is a high variation in the way OP is defined and, consequently, measured. In particular, a comprehensive analysis of the current OP practices as measured by Italian health care organizations is still missing. For this reason, our research aims at addressing the following research questions (RQs):

RQ1: How do Italian health care organizations define OP?

RQ2: Is OP measured by Italian health care organizations, and if so, how?

This chapter is organized into six sections: (1) literature review on OP with specific reference to the health care sector; (2) the Italian health care system; (3) methodology; (4) results; (5) discussion; and (6) conclusions.

2.2 Literature Review

Although managing performance is a wider concept than measuring it, performance management systems need to be based on sound measurement systems (Martin and Spano 2015). The performance measurement literature lacks consensus on concepts and definitions as well as on how OP may be measured (Au 1996; Forbes 1998; Ostroff 1992). Neely and Platts (1995, p. 9) comment that “performance measurement is a topic often discussed but rarely defined.” They also tried to provide a more specific definition of three concepts: performance measurement (“the process of quantifying the efficiency and effectiveness of action”); performance measure (“a metric used to quantify the efficiency and/or effectiveness of action”); and performance measurement system (“the set of metrics used to quantify both the efficiency and effectiveness of actions”).

The problem with these definitions is that they are too specific and, as a consequence, they do not convey what is now being labeled “performance measurement” in the literature and in practice (Bourne et al. 2003). In fact, over the past decades, performance was mainly measured only in its financial dimension

(McCracken et al. 2011) via “simple outcome-based financial indicators that are assumed to reflect the fulfillment of the economic goals of the firm” (Venkatram and Ramanujam 1986, p. 803). Most recent studies suggest a multidimensional approach to performance measurement considering the organization’s strategies as well (Nutti et al. 2013).

Since the advent of New Public Management (NPM) in the early 1990s, the issue of performance management has gathered increased attention (Bouckaert and Van Dooren 2009; Talbot 1999) and has become a fundamental issue for improving public services (Nutti et al. 2013). One of the reasons for this increased attention is the fact that governments started to be accountable for the use of public resources and for the results achieved (Bouckaert and Halligan 2008).

However, even though OP is particularly relevant to understanding why some organizations perform better than others, studies regarding OP in the public sector are not conclusive and there are different definitions of organizational performance (Andrews et al. 2011). In addition, several studies are based on perceived performance rather than on more objective measures, although there is evidence of a positive correlation between perceived OP and objective OP (Dollinger and Golden 1992). Both the reasons for measuring performance on the one hand and the process followed and the models used to measure it, on the other hand, are particularly important. With regards to the first aspect, Behn (2003) proposed eight purposes that public managers have for measuring performance: evaluate, control, budget, motivate, promote, celebrate, learn, and improve.

With regards to the second aspect, several authors describe how to design systems for performance measurement (Elg et al. 2013). For example, Kaplan and Norton (1992) consider that measures should be derived from strategy and represent different dimensions of an organization. Andrews et al. (2011) proposed a model for measuring OP in US federal agencies using three sets of measures: efficiency-related measures, effectiveness, and fairness. They identified five agency-level factors that may affect OP (organizational culture, human capital and capacity, agency support for the National Performance Review (NPR), leadership and supervision, and red tape) and four individual-level factors (structure of task/work, task motivation, public service motivation, and individual performance). They found that the most important elements that affect OP are effectiveness, teamwork, building human capital, structure of task/work, protection of employees, concern for the public interest, and task motivation. The main conclusion of this study is that OP is higher in organizations that adopt an involvement strategy—for both employees and other stakeholders.

In a similar Korean study, (Kim 2005) measured OP using a set of 12 items and provided evidence for the effect of individual-level factors on OP (such as job satisfaction, affective commitment, public service motivation, and organizational citizenship behavior). Kim (2005) investigated the link between OP and management innovation both directly and indirectly through performance management. In this study, OP was measured using a *core service performance score* constructed by the Audit Commission (2002) and based on six aspects of OP: quantity of outputs, quality of outputs, efficiency, formal effectiveness, equity, and consumer

satisfaction. Even though there is an established literature on this topic, the problem is that it remains a vaguely and loosely defined construct (Rogers and Wright 1998). In addition, several studies are based on perceived performance rather than on more objective measures, although there is evidence of a positive correlation between perceived OP and objective OP (Walker et al. 2011).

When discussing healthcare organizations, it is necessary to consider that they are complex adaptive systems (Anderson 1999; McDaniel et al. 2009) and, since the 1960s, complexity has been a central construct in the vocabulary of organization scientists (Anderson 1999). There are many ways in which this complexity can show itself (Daft 1992); however, even if the concept of complexity abounds in the public sector, the application of this theory is neither self-evident nor as straightforward as it might appear (Arnaboldi et al. 2015). In the specific case of health care organizations, the complexity relies on the phenomena's dynamism, which unfolds in unpredictable ways; these unfolding events are often unique, and it is interesting that a number of complexity theory advocates have identified health care as a suitable context for study (Arndt and Bigelow 2000). This complexity is also reflected in the way OP may be defined and measured. In fact, complexity theory has rich implications for the strategic management of organizations. Understanding this complexity to improve synergies among business units may improve OP. In the decades past, because of this complexity, measuring performance in the health care sector was uncommon and, in fact, it was believed that quality was not measurable. But today there is a higher interest in measuring and reporting performance in this sector, and in some cases there is the problem of having too many measures, some of which focus on outputs, outcomes, and processes, and others on single activities that have limited effect on overall health (Cassel et al. 2014).

Regarding the reasons for measuring performance in healthcare, according to De Vos et al. (2009), professionals use measurement for different purposes, i.e., evaluating, controlling, and improving clinical practice. Although there is little evidence that performance measures are actually used by practitioners to improve performance, Elg et al. (2013) suggest that "performance measurement may be a versatile method for driving improvement in healthcare organizations." In fact, performance measurement is recognized as a method with many utilization possibilities in health care (Elg et al. 2013). For example, implementing a transparent health care system is seen as a way to create external pressure and a sense of urgency for change (Elg et al. 2011). (Van der Wees et al. 2014) suggest that measures of quality are used by clinicians to evaluate the way they interact with patients and to measure quality improvement within their organizations; also, these measures may be used by health insurers to compare the performance of different providers. In addition, performance information may facilitate patients' decisions in choosing a provider.

Several studies have developed conceptual frameworks and models to help build effective OP measurement tools for the health care sector. For example, Arah et al. (2006) proposed a framework in which they present some common key performance dimensions for health care organizations. In building this framework, (Arah et al. 2006) considered other previous frameworks and the OP measurement systems used

in some jurisdictions (UK, Canada, Australia, USA, European Community Health Indicators, World Health Organization, and OECD) and created a list of performance dimensions in healthcare: effectiveness, appropriateness, safety, efficiency, continuity, accessibility, equity, responsiveness, patient-centeredness, timeliness, and acceptability. Some of these dimensions are consistent with the dimensions required by the Italian legislation, even if defined in a different way.

A recent study proposed a new model for measuring and evaluating health care organizations' OP using two main dimensions: outcome and delivery efficiency. The model is based on a "matrix of performance evaluation" (Elg et al. 2011) and includes 42 indicators, 24 concerning outcome and 18 on efficiency, and an additional area related to "management."

Studies on OP in Italian health care organizations are limited. Baraldi and Bocci (2009) analyzed the most common methodologies to measure OP of Italian health care organizations. In particular, they surveyed how Italian health care organizations measure their performance and observed the increased importance of the balanced scorecard that has been adapted to the features of the health care sector. In fact, even though financial indicators are still used—as in profit-oriented organizations—many nonfinancial indicators have taken center stage, and the balanced scorecard is useful to measure both financial and nonfinancial performance in health care organizations (Nuti et al. 2013). Bocci (2005) proposed a new model of the balanced scorecard for health care organizations based on four perspectives (community, internal process, financial resources, and learning and growth).

In 2005, the *Istituto Superiore Sant'Anna* of Pisa created and implemented in some health organizations in Tuscany a new OP measurement method based on the balanced scorecard model. This method is based on six evaluation dimensions (population health status, capacity to pursue regional strategies, clinical performance, patient satisfaction, staff satisfaction, efficiency, and financial performance) (Nuti et al. 2013). For each dimension, a set of indicators is defined ~ 130 indicators; the balanced scorecard approach is then used to evaluate OP. Since 2005, this method has been introduced in other organizations in Italy. In particular, eight other Italian regions and the Ministry of Health have adopted the S. Anna method to monitor levels of health services provided in the country (Nuti et al. 2013). This system is, as can be seen by the above description, a multidimensional performance measurement system and has been valued as particularly innovative and comprehensive.

However, as highlighted by Baraldi and Bocci (2009), the most common performance measurement methodologies in health care organizations are budgeting, cost accounting, and accounting for responsibility centers. These results show that Italian public healthcare organizations mainly focus on OP's financial dimension.

Broadly speaking, the OP literature in the Italian health care sector is limited, and there are few analyses of the actual measurement and evaluation systems. To fill this gap, this chapter focuses on organizational performance and concentrates on the Italian health care sector by addressing the following research questions:

RQ1: How do Italian health care organizations define OP?

RQ2: Is OP measured by Italian health care organizations, and if so, how?

2.3 The Italian Health care System

Italy's healthcare system (*Servizio Sanitario Nazionale* [SSN]) provides universal coverage free of charge at the point of service. The system is organized into three levels: national, regional, and local (Lo Scalzo et al. 2009). The general objectives and the fundamental principles of the health care system are guaranteed by the national level, while services are delivered at the regional level through local health organizations (Van der Wees et al. 2013) and public and private hospitals.

This system is based on public financing via general taxation. There are also private health organizations that provide health services. In particular, the percentage of hospital beds supplied by public sector organizations is 80.7%, with the remainder supplied by nonprofit and private organizations (Trincherio et al. 2013). The organizations that provide health care services are as follows:

- Local health authorities (*Aziende Sanitarie Locali*, ASLs)
- Public hospitals (*Aziende Ospedaliere*, AOs)
- Research Institutes for Hospitalization and Medical Treatment (*Istituto di ricovero e cura a carattere scientifico*, IRCCSs)
- Private accredited providers

The local health authorities are responsible for providing a selection of health services. Each region may have many ASLs, with each ASL responsible for providing healthcare to a given population. Initially, there were 659 ASLs, but several reforms occurred in the 1990s to modify their function and governance system. Their number was further reduced in 2015 to 139. The ASLs provide care directly through their own facilities and also buy services from external suppliers such as accredited private providers.

Public hospitals, established by Legislative Decree No. 502/1992 and defined as quasi-independent agencies, enjoy financial and operating autonomy. In 1995, many preexisting hospitals were transformed into 82 AOs. This was further reduced to 77 in 2015. There are three necessary conditions to obtain AO status: "a divisional organizational structure; the existence of at least three clinical units; and a complete emergency department with an intensive care unit" (Lo Scalzo et al. 2009, p. 76). AOs provide healthcare to all residents in a region, while ASLs serve a portion of the population. Also, AOs are financed based on the diagnosis-related group (DRG) system, while ASLs are financed based on per capita transfers.

The IRCCSs are research-oriented hospitals operating at the local level with competences in research and treatment of important diseases. In 2008, 13 of the 20 Italian regions had 42 IRCCSs divided into 18 public and 24 private institutions. As of 2015, there are 21 public and 27 private IRCCSs in Italy. The scientific activities of the hospitals are monitored by the Ministry of Health, which is also responsible for establishing new IRCCS.

Since 1990, Italy's health care system has seen several reforms introduced by different pieces of legislation (Law N. 833/1978, Legislative Decrees N. 502/1992, N. 517/1993, and N. 229/1999) that have changed its structure and established the procedures now in use.

With regards to the issue of OP in the Italian health care system, Legislative Decree 150 of 2009 introduced the following eight dimensions:

1. Implementation of active policies for satisfying citizens' needs;
2. Implementation of plans and programs;
3. Customer satisfaction;
4. Modernization and qualitative improvement of public organizations and employees' professional skills and the capability to implement plans and programs;
5. Improvement of relations with citizens and other stakeholders;
6. Efficiency in the use of resources, with particular reference to cost reduction;
7. Quality and quantity of services; and
8. Equal opportunities.

Our analysis focused on the effective use of these and additional dimensions of OP by health organizations.

2.4 Methodology

The research looks at the way organizational performance is (1) defined and (2) measured by Italian healthcare organizations. The data collection methods include document analysis and semi-structured interviews with key informants. To investigate the ways in which the healthcare organizations define OP and measure it, we performed an in-depth analysis of the content of the documents prepared by a sample of Italian public health care organizations. In addition, we analyzed the performance documents of the seven Italian health care organizations that are accredited by the Joint Commission—an independent, not-for-profit organization that accredits and certifies top performing health care organizations and programs in the USA and across the world (Joint Commission International 2016). In Italy, there are seven accredited public health organizations:

1. AO *Santa Maria degli Angeli*;
2. ASL 3 *Alto Friuli*;
3. AOU *Santa Maria della Misericordia*;
4. *Ospedale Cattinara*;
5. *Istituto Giannina Gaslini*;
6. *Presidio Ospedaliero Oglio-Po*; and
7. *Ospedale Santa Chiara*.

Content analysis is a research method that “classifies textual material, reducing it to more relevant, manageable bits of data” (Weber 1990, p. 5). In particular, we used an inductive approach, starting with data and then creating specific categories that can explain the general phenomena. The qualitative data were organized with the process of “open coding” according to which notes and headings were written in the text while reading it. Only after this analysis was the categories created.

The analyzed documents include the following: the evaluation system, the performance plan, and the performance report. Each of these documents has specific functions, and it is important to consider all of them in the analysis. The evaluation system sets the guidelines by which performance at both individual and organizational levels is measured and evaluated. The performance plan shows what performance dimensions, objectives, and indicators have been selected, consistent with the evaluation model defined by the system. The performance report provides evidence of the results achieved and of the way the performance measurement process worked. These are the specific documents requested by the legislation on performance management in Italian public organizations (Legislative Decree 150/09).

The census of Italian public health care organizations is composed of 237 units. These organizations are divided into 139 ASLs, 77 AOs, and 21 IRCCSs. For analysis, a random sample of 20% was extracted via stratified samples. In this way, the study was conducted through a sample of 50 health care organizations and was subdivided in 30 local health authorities (ASLs), 16 public hospitals (AOs), and four research institutes (IRCCS). A set of substitutes was randomly extracted as alternatives. During the first step of the extraction, we replaced some selected organizations that had not published their performance plans on their Web sites. These included nine healthcare organizations (18% of the overall sample) that had not published performance plans and were subdivided in four local health authorities (13% of the 30 extracted authorities), four public hospitals (25% of the 16 selected hospitals), and one public National Institute for Scientific Research (25% of the four selected institutes). These organizations have been replaced with other organizations that did publish a performance plan. This way, the sample is composed only of organizations with officially published performance plans.

To answer the first research question, we noted in each document whether and how OP is defined. We also clustered the definitions to identify recurrent aspects and which organizations comply and do not comply with the legislation. We also searched for innovative ways to define and measure OP.

To answer the second research question, we studied the measurement systems regarding OP, focusing on both methodological and practical aspects. This analysis was made among the ASLs, AOs, and IRCCSs. In addition, all performance reports were clustered using three criteria: strategic areas, objectives, and performance dimensions.

We also identified congruence among the three different analyzed documents. In particular, the study focused on the performance dimensions used in the measurement process. During the analysis of their congruence, we considered whether, in every document, the same performance dimensions were reported. Broadly speaking, we studied whether each document fulfilled its tasks.

To strengthen the results of the document analysis, 30 qualitative semi-structured interviews were conducted between May and August 2016. Two general directors, three administrative directors, and 25 organization and control managers were interviewed. The interviews lasted about 40 min and were recorded and transcribed. With regards to the regional distribution, nine interviewees belong to organizations that are located in the northwest of Italy, 10 in the northeast, three in the center, and

eight in the south and the islands. The interviewees were asked to answer questions related to the performance management systems used in their organizations and were also allowed to add other comments about the specific performance dimensions measured. The interviews provided a deeper understanding of the ways in which Italian healthcare organizations effectively measure their OP, strengthening the results of the document analysis or, in some cases, highlighting the differences.

2.5 Results

Our research revealed that just 34% of the organizations (ranging from 25% of AOs and IRCCSs to 33.3% of ASLs) published the evaluation system and a 78% published the performance reports on their Web sites (ranging from 50% of IRCCSs to 83.3% of ASLs) (see Table 2.1). If we consider the initial extraction of the sample, before the substitutions, 18% of the organizations did not publish the performance plan (13% of the local health authorities, 25% of the public hospitals, and 25% of National Institutes for Scientific Research). This means that just 87, 75, and 75%, respectively, published the performance plan (Table 2.1).

The first RQ describes how OP is defined by Italian health care organizations. The results show that there are many differences among Italian public health organizations in the way OP is defined and measured. In addition, not all organizations explicitly provided a definition of OP. In particular, 62% did not provide any definition at all (57% of ASLs, 69% of AOs, and 75% of IRCCS). The remaining 38% of the organizations explicitly defined OP. Of the organizations providing a definition, 79% (15 out of 19 organizations) used the very same definition provided by the legislation (60% of ASLs, 60% of AOs, and 100% of IRCCS): “The contribution that a subject generates through its action to achieve the purposes and the objectives, and to satisfy the needs for which the organization has been created” (Delibera Civit 89/2010). In three cases only, different definitions were chosen. For instance, one organization defined OP as “the performance obtained by the firm as a whole and by each organizational unit.” In some cases, even if there is not a specific definition of the performance dimensions, the definition itself has been derived from the strategic areas as defined in the performance plan (this is true for 6 out of 30 ASLs and 2 out of 16 AOs) or from the objectives (3 out of 30 ASLs). The interviews confirmed these results. In fact, most interviewees did not provide an explicit definition and told us that no specific dimensions are used to measure OP. Respondents reported the way that OP was

Table 2.1 Published performance documents

	Evaluation system (%)	Performance plan (%)	Performance report (%)
ASL	33.3	87	83.3
AO	25	75	75
IRCCS	25	75	50

measured, regardless of its definition. When a definition was given, it was the same as the legislation. In some cases, strategic areas are defined in a way that is consistent with the OP dimensions as defined by the legislation. For this reason, the performance plans of the organizations in the sample were analyzed and contrasted according to two elements—strategic areas and objectives—to understand the underlying performance dimensions.

With respect to whether and how OP is measured by Italian healthcare organizations, after analyzing the performance documents, we then focused on specific performance dimensions, and we contrasted the dimensions used by the organizations with the eight dimensions provided by the legislation (Article 8 of Legislative Decree 150/2009). By analyzing all published documents (evaluation system, performance plan, and performance report), we verified the specific dimensions that health organizations actually use to measure and evaluate OP (Table 2.2). This analysis shows that the evaluation systems report just a minority of the eight dimensions of OP introduced by the legislation listed above. They range from 50% of the cases for “quality and quantity of services delivered” to 0% of the “qualitative and quantitative development of relationships with the relevant stakeholders” (see Table 2.2; Annex 1). Only 56% of the organizations specified the performance dimensions used in the measurement process in their performance plan (11 ASLs, 13 AOs, and 4 IRCCSs). In the performance report, the presence of the OP dimensions ranges from 64% for “efficiency in the use of resources” to 15% for “equal opportunities.” The performance plans show the highest percentage of the presence for all the dimensions with a range from 78% for “efficiency in the use of resources” to 20% for “equal opportunities.”

The most recurrent OP dimensions are “efficiency in the use of resources” and “quality and quantity of delivered services” (Table 2.3).

In just one case, OP was actually defined and measured using all eight dimensions provided by the legislation (as emerged from both the performance plan and the performance report). The other organizations measured only some of the dimensions requested by the legislation. In almost 60% of cases, the organizations introduced additional dimensions not required by the legislation. In particular, the most recurring performance dimensions in the performance report that differ from the legislation are appropriateness, risk management, processes, research, and teaching (Table 2.4).

Appropriateness is divided into two elements: clinical appropriateness and organizational appropriateness. “Clinical appropriateness” applies to cases in which

Table 2.2 % of OP dimensions present in the performance documents

	Performance dimensions							
	1	2	3	4	5	6	7	8
ES (%)	25	13	31	31	0	38	50	6
PP (%)	48	30	36	44	26	78	74	20
PR (%)	44	21	28	38	23	64	62	15

ES Evaluation System, PP Performance Plan, PR Performance Report

Table 2.3 Most recurrent OP dimensions

Type of perf. document	Efficiency				Quality and quantity			
	ASL (%)	AO (%)	IRCCS	Overall (%)	ASL (%)	AO (%)	IRCCS	Overall (%)
ES	13.3	12.5	–	12	16.7	18.8	–	16
PP	73.33	81.25	100%	78	66.7	81.3	100%	74
PR	47	69	50%	50	43	56	50%	48

ES Evaluation System, PP Performance Plan, PR Performance Report

Table 2.4 OP dimensions present in the performance report not listed in the legislation

Performance dimensions	ASL	AO (%)	IRCCS
Appropriateness	30%	44	0
Research and teaching	7%	6	75%
Risk management	1%	12	
Processes	0	31	0

healthcare interventions occur in such conditions that the probability of benefits outweighs the potential risks in terms of safety for the patient and economy of resources (Scaletti 2014). “Organizational appropriateness” refers to the fact that health care interventions must be provided at the proper level of assistance. For instance, patients that may be treated in a day hospital center should not be admitted to hospitals. Most organizations use organizational appropriateness rather than clinical appropriateness as a measure.

Focusing on the performance report (which shows what is actually done in terms of performance measurement and evaluation), 22% of the surveyed health organizations did not report any OP dimensions. In fact, 11 out of the 50 surveyed organizations do not mention OP measurements in their performance report at all. In another 18% of cases, the OP measurement is limited to a small number of dimensions. As a consequence, just 60% of the organizations in our sample make some sort of OP measurement using one or more of the eight performance dimensions required by the legislation.

2.6 Discussion

Measuring and reporting performance in health care is recognized as an important tool to improve the quality of the services delivered by health care organizations (De Vos et al. 2009; Elg et al. 2013). However, the actual use of performance measurements in the health care sector is also limited because of a lack of understanding of how these measures must be used in practice (Elg et al. 2013). More generally, performance information allows governments to monitor health care systems’ performance (De Vos et al. 2009).

The results obtained in the research raise some areas of concern. First, we found that just a limited percentage of the organizations (34%) published the evaluation system on their Web sites. Although it is not compulsory under the current legislation, we believe that such reporting of the methods for evaluation is not fully consistent with the principle of transparency that aims to provide citizens and other stakeholders with all the relevant information needed to learn the results achieved and hold these organizations accountable.

Second, the majority of organizations (62%) did not provide any definition at all of OP and those that did, used the very same definition provided by the legislation, which is very general and even vague. A lack of clarity in the way OP is defined does not help in understanding the actual results achieved and does not give a sense of direction to the people working in the organizations.

Third, the analysis of the performance documents highlighted that Italian public health organizations are only partially complying with the legislation. For example, there is a difference between what is said in the evaluation system, in the performance plan, and in the performance report regarding the measurement of OP. In fact, the evaluation systems and the performance plans promise more than the performance reports deliver. These data may be explained by considering that health organizations have set the evaluation systems in a very generic way and have used the performance plan to better specify the content of the OP dimensions and how to measure them. To some extent, it seems that they tend to underestimate the difficulty in measuring and evaluating OP. As a consequence, when it comes to reporting what dimensions of OP have actually been measured and evaluated, the reported percentages are lower. Fourth, public health organizations did not find in the legislation a model that fits with the peculiar features of the healthcare sector. We found that 58% of the sample uses dimensions of OP different from those in the legislation—mainly appropriateness and risk management. In particular, appropriateness is particularly relevant in the health care sector. The interviews showed that appropriateness is a dimension used by all organizations to which interviewees belong, but it is reported in only one-third of the analyzed documents. Given the very nature of the health care sector, risk management is also very important—as demonstrated by its presence among the performance dimensions.

Not surprisingly, IRCCSs make extensive use of the research and teaching dimensions, given their specificity. In fact, IRCCSs are research-oriented hospitals in which research and teaching dimensions are fundamental.

Broadly speaking, it appears that the actual measurement and evaluation of OP by Italian health organizations is limited, and those that actually perform it use only a limited number of performance dimensions. In addition, there is significant variability in the content of OP and in the process of measuring and evaluating it.

To have a clearer picture of the Italian health care situation, we analyzed the performance documents of the Italian public health organizations that are accredited by the Joint Commission (Joint Commission International 2016). The Joint Commission's accreditation is a guarantee of quality of the health care services provided by the accredited organizations. In fact, the validation process is based on international standards of excellence in performance and organization to guarantee

security and high quality of the services. The analysis of the content of the performance documents of the organizations accredited by the Joint Commission shows a similar situation compared to the sample. In fact, all the accredited organizations published their performance plans on their Web sites. Six out of the seven published their performance reports, but only two (30%) published the evaluation system. Even for the most advanced public health organizations, the importance of publishing the evaluation system is apparently low. We compared the performance dimensions required by the legislation with the performance dimensions actually used by the accredited organizations. The analysis shows some differences with regards to the most used OP dimensions versus the sample. In fact, all accredited organizations consider in their performance plans two specific dimensions: the “implementation of plans and programs” and the “modernization and qualitative improvement of public organizations and employees’ professional skills and the capability to implement plans and programs.” In the organizations studied here, we found that the most commonly used performance dimensions are “efficiency” and “quality and quantity of services.” The additional dimensions used by the accredited organizations are the same cited previously (appropriateness, risk management, and research). Some of the interviewees are from organizations accredited by the Joint Commission. What emerged is that the only difference in comparison to the nonaccredited organizations is a higher attention to the quality of the performance:

Some objectives are connected to the quality of the performance because the Joint Commission asks us to maintain and to show specific standards of quality. (Interviewee 17)

In fact, the accreditation program requires some qualitative parameters to be met, so the healthcare organizations must measure these aspects with more attention than others to make sure that they meet the required levels of quality.

The semi-structured interviews showed some other interesting results. All interviewees recognized the importance and the relevance of the performance measurement system in place. All of them said that having a good performance measurement system is a necessary condition to effectively manage their organization—particularly with regards to complex organizations like those in the health care sector. The interviews showed that in most organizations, the performance measurement system is not sufficiently embedded into the organizational structure and is continually being changed and improved over time:

If I look at the first performance plan, it looks really embryonic; but year after year we improved it. Maybe if I look at the present performance plan in three years, I will realize it has been done in the wrong way. (Interviewee 19)

In particular, the first relevant issue is about the role of regional legislation in defining the performance objectives of the health care organizations. Broadly speaking, each regional government defines some performance objectives that have to be achieved by every health care organization in that region. Thus, the starting point of every performance measurement system is the regional legislation. Another interesting common element is the role of the director general and his influence in structuring the performance measurement system and its operation.

The interviewees highlighted that the presence of a director general sensitive to the issue of performance measurement positively influences the effectiveness of the performance system itself, as reported by one interviewee:

In this moment the top management focuses only on financial aspects, and I am sorry about it, because with the previous director general the OP measurement system was more complete. (Interviewee 2)

Broadly speaking, if the director general pays attention to the ways in which OP is measured, then the organization as a whole is more likely to have a more effective performance measurement system; on the contrary, if the director general does not care about this issue, then measuring OP will be neglected with negative consequences for the organization as a whole. In two cases only, the interviewees reported that the system was already well structured when a new director general not attentive to OP measurement came in. This did not hamper the functioning of the systems itself.

2.7 Conclusions

The healthcare sector in Italy has traditionally been at the forefront of the innovations and reforms that took place in this country. For example, in the early 1990s, health organizations were the first to introduce accrual accounting and management tools.

Although OP is a particularly relevant topic, it is still neglected in the Italian public sector, which has focused more on individual performance than on organizational performance (Agasisti and Arnaboldi 2011). The Italian healthcare sector is not different, and often neglects OP. In fact, our analysis shows that 40% of organizations in our sample do not appear to measure and evaluate OP. The remaining 60% undertake some form of measurement and evaluation of performance at organizational level.

Our research provides evidence of a significant variance in the way OP is defined and measured, with subsequent comparison problems. In some cases, this variance may be, at least in part, explained by the different types of organizations, i.e., local health authorities are different from public hospitals and from research institutes for hospitalization and medical treatment. However, there is also a significant variance among organizations of the same kind. Although this difference is, to some extent, normal, it shows a limited alignment of performance measurement systems among Italian health organizations. It also shows the difficulty deriving from implementing a top-down performance management system enforced by law (Micheli and Neely 2010). In addition, the research confirms the persistence of two traditional problems of the Italian public sector. One refers to the limited attention given to the role and importance of performance management (Martin and Spano 2015). The second is the false conviction that changes can be introduced by law.

Our study tried to fill the gap in the literature related to OP in the Italian health care sector by providing an in-depth analysis of the way OP is defined and measured. Some implications emerge from this research. First, the cited limited

compliance with the legislation, in a country where formal respect of the legislation is paramount (Martin and Spano 2015), needs to be better analyzed and understood. In fact, the strong presence of OP dimensions that differ from those listed in the legislation confirms one of the most criticized aspects of the existing legislation, which is that the legislation is the same for every kind of public organization regardless of differences in typology, size, specific context, and the like (Giovanelli et al. 2015). For example, none of the eight cited dimensions is specifically suitable for the health care sector. Thus, it is not surprising that a significant portion of the organizations in this sector decided to complement the legislation with other dimensions that are perceived to be more useful in capturing what OP is in this specific context. In fact, the legislation sets the general rules that are the same for all public organizations in Italy, regardless of the many existing differences among the different types of organizations, e.g., municipalities and health organizations. This emphasizes the need to adapt the set of OP dimensions prescribed by the legislation to the specific context. Thus, it is no surprise that in the case of the healthcare sector, some organizations select different dimensions from the ones provided by the legislation. Consequently, the overall framework that imposes the same rules to all Italian public authorities and agencies needs to be revisited.

Second, our data show that there is limited actual use of performance measurement by Italian public health care organizations, and a significant portion of the surveyed organizations do not measure OP. The limited attention to the definition of OP and its measurement has been confirmed, to some extent, by the analysis of the health care organizations accredited by the Joint Commission, i.e., those organizations that should represent the best practices in terms of organizational performance. Nevertheless, even accredited organizations do not measure all the performance dimensions required by the legislation; they measure just a portion. While the most common OP dimensions of the organizations in the sample are “efficiency” and “quality and quantity of services,” accredited organizations more often use two other dimensions: “implementation of plans and programs” and “modernization and qualitative improvement of public organizations and employees’ professional skills and the capability to implement plans and programs.” While it is no surprise that quality improvement is of paramount importance for accredited organizations, it would be interesting to better understand the reasons underlying the different importance given to the other OP dimensions. In addition, accredited organizations use the same additional dimensions introduced by the other organizations in the sample, such as “appropriateness” and “risk management.” This confirms on the one hand that these two elements are very important in the health care sector, and, on the other hand, that there is a need for a general cultural change to foster a stronger commitment to measuring and evaluating performance and a realization that a centralized approach is not always the best choice (Micheli and Neely 2010).

Third, there appears to be a shortfall in designing and implementing performance management policies and frameworks that are homogenous across the Italian regions and that flow from the national to the regional and local levels, creating what are considered 20 different health care systems (one for every Italian region) (Bertin and Cipolla 2013).

References

- Agasisti, T., & Arnaboldi, M. (2011). Debate: Public sector productivity: lessons from the Italian financial crisis. *Public Money and Management*, 31(6), 383–384. doi:10.1080/09540962.2011.618758.
- Anderson, P. A. (1999). Complexity theory and organization science. *Organization Science*, 10(3), 216–232.
- Andrews, R., Boyne, G. A., & Walker, R. M. (2011). Dimensions of publicness and organizational performance: A review of the evidence. *Journal of Public Administration Research and Theory*, 21(Supplement 3), i301–i319. doi:10.1093/jopart/mur026.
- Arah, O. A., Westert, G. P., Hurst, J., & Klazinga, N. S. (2006). A conceptual framework for the OECD health care quality indicators project. *International Journal for Quality in Health Care*, 18(Supplement 1), 5–13.
- Arnaboldi, M., Lapsley, I., & Steccolini, I. (2015). Performance management in the public sector: The ultimate challenge. *Financial Accountability and Management*, 31(1), 1–23.
- Arndt, M., & Bigelow, B. (2000). Commentary: The potential of chaos theory and complexity theory for health services management. *Health Care Management Review*, 25(1), 35–38.
- Au, C. (1996). Rethinking organizational effectiveness: Theoretical and methodological issues in the study of organizational effectiveness for social welfare organizations. *Administration in Social Work*, 20, 1–20.
- Audit Commission. (2002). *Comprehensive performance assessment*. Retrieved from London.
- Baraldi, S., & Bocci, F. (2009). Il Balanced Scorecard nelle aziende sanitarie: 5 anni dopo. Università Cattolica del Sacro Cuore.
- Behn, R. D. (2003). Why measure performance? different purposes require different measures. *Public Administration Review*, 63(5), 586–606.
- Bertin, G., & Cipolla, C. (Eds.). (2013). Verso differenti sistemi sanitari regionali: Edizioni Ca'Foscari.
- Bevan, G., & Hood, C. (2006). What's measured is what matters: Targets and gaming in the English public health care system. *Public Administration*, 84(3), 538–571. doi:10.1111/j.1467-9299.2006.00600.x.
- Bocci, F. (2005). A multi-dimensional approach to the community perspective in the balanced scorecard architecture for the public sector. Empirical evidence in healthcare organizations. EGPA 2005 Conference.
- Bouckaert, G., & Halligan, J. (2008). *Managing performance. International comparisons*. New York: Routledge.
- Bouckaert, G., & Van Dooren, W. (2009). Performance measurement and management in public sector organizations. In Tony Bovaird & Elke Löffler (Eds.), *Public management and governance* (2nd ed., pp. 151–164). Abingdon: Routledge.
- Bourne, M., Mills, J., Wilcox, M., Neely, A., & Platts, K. (2003). Designing, implementing and updating performance measurement systems. *International Journal of Operations and Production Management*, 20(7), 754–771.
- Cassel, C. K., Conway, P. H., Delbanco, S. F., Jha, A. K., Saunders, R. S., & Lee, T. H. (2014). Getting more performance from performance measurement. *The New England Journal of Medicine*, 23(371), 2145–2147.
- Daft, R. L. (1992). *Organization Theory and Design*. Mason: South-Western, Cengage Learning.
- De Vos, M., Graafmans, W., Kooistra, M., Meijboom, B. R., Van Der Voort, P., & Westert, G. P. (2009). Using quality indicators to improve hospital care: A review of the literature. *International Journal for Quality in Health Care*, 21(2), 119–129. doi:10.1093/intqhc/mzn059.
- Dollinger, M. J., & Golden, P. A. (1992). Interorganizational and collective strategies in small firms: Environmental effects and performance. *Journal of Management*, 18(4), 695–715. doi:10.1177/014920639201800406.

- Elg, M., Palmberg, K., & Kollberg, B. (2013). Performance measurement to drive improvements in healthcare practice. *International Journal of Operations and Production Management*, 33(11/12), 1623–1651. doi:10.1108/IJOPM-07-2010-0208.
- Elg, M., Stenberg, J., Kammerlind, P., Tullberg, S., & Olsson, J. (2011). Swedish healthcare management practices and quality improvement work—development trends. *International Journal of Health Care Quality Assurance*, 24(2), 101–123. doi:10.1108/095268611111105077.
- Forbes, D. P. (1998). Measuring the unmeasurable: Empirical studies of nonprofit organization effectiveness from 1977 to 1997. *Nonprofit and Voluntary Sector Quarterly*, 27, 183–202.
- Giovanelli, L., Marinò, L., Rotondo, F., Fadda, N., Ezza, A., & Amadori, M. (2015). Developing a performance evaluation system for the Italian public healthcare sector. *Public Money and Management*, 35(4), 297–302. doi:10.1080/09540962.2015.1047274.
- Joint Commission International. (2016). <http://www.jointcommissioninternational.org>.
- Kaplan, R. S., & Norton, D. P. (1992). The balanced scorecard—measure that drive performance. *Harvard Business Review*, 70(1), 71–79.
- Kim, S. (2005). Individual-level factors and organizational performance in government organizations. *Journal of Public Administration Research and Theory*, 15(2), 245.
- Lo Scalzo, A., Donatini, A., Orzella, L., Cicchetti, A., Profili, S., & Maresso, A. (2009). Italy: Health system review. *Health Systems in Transition*, 11(6), 1–216.
- Martin, J., & Spano, A. (2015). From performance management to strategic local government leadership: Lessons from different cultural settings. *Public Money and Management*, 35(4), 303–310.
- McCracken, M. J., McIlwain, T. F., & Fottler, M. D. (2011). Measuring organizational performance. *Health Services Management Research*, 14(4), 211–219. doi:10.1258/0951484011912717.
- McDaniel, R., Lanham, H. J., & Anderson, R. A. (2009). Implications of complex adaptive systems theory for design of research on health care organizations. *Health Care Management Review*, 34(2), 191–199.
- Micheli, P., & Neely, A. (2010). *Performance measurement in the public sector in England: Searching for the golden thread*, 70(4), 591–600. doi:10.1111/j.1540-6210.2010.02180.x.
- Neely, A., & Platts, M. G. K. (1995). Performance measurement system design. A literature review agenda. *International Journal of Operations and Production Management*, 15(4), 80–116.
- Nuti, S., Seghieri, C., & Vainieri, M. (2013). Assessing the effectiveness of a performance evaluation system in the public health care sector: Some novel evidence from the Tuscany region experience. *Journal of Management and Governance*, 17(1), 59–69.
- Ostroff, C. (1992). The relationship between satisfaction, attitudes and performance. An organizational level analysis. *Journal of Applied Psychology*, 6(77), 963–974.
- Pollitt, C. (2005). Performance management in practice: A comparative study of executive agencies. *Journal of Public Administration Research and Theory*, 6(1), 25–44. doi:10.1093/jopart/mui045.
- Rogers, E. W., & Wright, P. M. (1998). Measuring organizational performance in strategic human resource management: Problems and prospects (CAHRS Working Paper #98-09). Ithaca, NY: Cornell University, School of Industrial and Labor Relations, Center for Advanced Human Resource Studies. <http://digitalcommons.ilr.cornell.edu/cahrswp/122>.
- Scaletti, A. (2014). *Evaluating Investments in Health Care Systems*. New York: Springer International Publishing.
- Smith, P.C., Mossialos, E., & Papanicolas, I. (2008). *Performance measurement for health system improvement: Experiences, challenges and prospects*. Retrieved from <http://www.who.int/management/district/performance/PerformanceMeasurementHealthSystemImprovement2.pdf>.
- Talbot, C. (1999). Public performance—towards a new model? *Public Policy and Administration*, 14(3), 15–34. doi:10.1177/095207679901400302.
- Trincherò, E., Brunetto, Y., & Borgonovi, E. (2013). Examining the antecedents of engaged nurses in Italy: Perceived organisational support (POS); satisfaction with training and development; discretionary power. *Journal of Nursing Management*, 21, 805–816. doi:10.1111/jonm.12143.

- Van der Wees, P. J., Zaslavsky, A. M., & Ayanian, J. Z. (2013). Improvements in health status after Massachusetts health care reform. *Milbank Quarterly*, *91*(4), 663–689. doi:[10.1111/1468-0009.12029](https://doi.org/10.1111/1468-0009.12029).
- Van der Wees, P. J., der Nijhuis-van Sanden, M. W. G., van Gineken, E., Ayanian, J. Z., Schneider, E. C., & Westert, G. P. (2014). Governing healthcare through performance measurement in Massachusetts and the Netherlands. *Health Policy*, *116*(1), 18–26. doi:[10.1016/j.healthpol.2013.09.009](https://doi.org/10.1016/j.healthpol.2013.09.009).
- Venkatram, M., & Ramanujam, V. (1986). Measurement of business performance in strategy research: A comparison of approaches. *Academy of Management Review*, *1*(4), 801–814.
- Walker, R. M., Damanpour, F., & Devece, C. A. (2011). Management innovation and organizational performance: The mediating effect of performance management. *Journal of Public Administration Research and Theory*, *21*(2), 367–386. doi:[10.1093/jopart/muq043](https://doi.org/10.1093/jopart/muq043).
- Weber, R. T. (1990). *Basic content analysis*. Newbury Park, CA: Sage.

Chapter 3

The Iron Law of Unintended Effects, Again? Outcome Measures and Blame-Avoidance

Andrea Garlatti, Paolo Fedele and Mario Ianniello

Abstract The shift from output to outcome measures is a recurrent doctrine in public administration studies and practice. However, as with many popular doctrines before, more empirical analysis is still needed. This chapter focuses on the unintended effects of outcome-based performance management and explores how the use of outcome measures influences blame-avoidance strategies by officials and service providers. In looking for answers and using the concept of social mechanism as the analytical lens, this contribution explores a pilot case in the Italian public sector, where a performance ranking composed of outcome measures was introduced as the pivotal performance management tool. Results allow to conceptualize a link between the type of blame-avoidance response and the features of the potential blamers.

Keywords Outcome measures · Blame-avoidance · Social mechanisms · Exploratory case study

3.1 Introduction and Research Question

The shift from output to outcome measures is a recurrent doctrine in public administration and management studies (Massaro et al. 2015; Van Dooren et al. 2010; Bouckaert and Halligan 2008; Boyne et al. 2007). However, as with many popular management doctrines before, more empirical analysis could prove fruitful. In this light, this chapter adopts a specific angle: It looks into the unintended effects (Hood and Peters 2004; Maor 1999; Sieber 1981; Merton 1936) of using outcome measures.

A. Garlatti (✉) · P. Fedele · M. Ianniello
University of Udine, Udine, Italy
e-mail: andrea.garlatti@uniud.it

P. Fedele
e-mail: paolo.fedele@uniud.it

M. Ianniello
e-mail: mario.ianniello@uniud.it

Specifically, it analyzes one potential cluster of unintended effects, i.e., the blame-avoidance strategies (Hood et al. 2016; Hood 2007b, 2013; Weaver 1986, 1987) that officials and service providers could enact when outcome measures are used. Therefore, it addresses, in an exploratory fashion, the following research question: How does the use of outcome measures influence blame-avoidance strategies? In looking for answers, this contribution explores a pilot case in the Italian public sector, where a performance ranking composed of outcome measures was introduced as the pivotal performance management tool, so to insulate potential mechanisms (Hedstrom 2005) that lead to the enactment of blame-avoidance.

This chapter adds to current knowledge in different ways. Although the analysis of performance paradoxes has a quite long history (Adcroft and Willis 2005; Van Thiel and Leeuw 2002; Bouckaert and Balk 1991), more empirical studies are surely needed. Even though typologies of performance paradoxes and unintended effects have been elaborated and some studies have performed empirical observations (Bevan and Hood 2006), the very existence and nature of the unexpected developments associated with performance management is still controversial (Gerrish 2016; Kelman and Friedman 2009; Wenger et al. 2008). This chapter aims at contributing to this debate and, in addition, displays some relative novelties. Firstly, most previous contributions (with some exception, see Grizzle 2002) looked at the paradoxes induced by performance measures as such, while this chapter focuses solely on outcome measures. Secondly, the present analysis looks at a specific form of unintended effect, namely blame-avoidance strategies. Although few previous contributions analyzed the relation between performance information and blame-avoidance (Nielsen and Moynihan 2016; Charbonneau and Bellavance 2012), the latter is still under-analyzed, at least as a form of performance paradox. Thirdly, while most analyses of performance paradox focused on target systems, this chapter focuses on rankings instead (Arndt 2008; Hood 2007a, Van de Walle 2008).

Results can contribute, on the one hand, to performance management research in the public sector, adding new empirical evidence on one specific angle of a well-known phenomenon. Furthermore, this chapter can contribute to public management research more in general, especially to the research agenda aimed at exploring its paradoxes and unintended effects. Finally, on a more practical side, findings can provide decision makers with insights about how to avoid some traps in designing and using outcome measures.

3.2 Theoretical Background

3.2.1 *Public Management Paradoxes: Picking a Theoretical Lens*

The analysis of the paradoxes (Maor 1999; Sieber 1981; Brams 1976) associated with public management has become increasingly popular with the “middle-aging” of NPM (Hood and Peters 2004). Numbers of scholars have moved from

descriptive accounts of management practices to finer grain analysis of developments that are unexpected or contrary to received beliefs. Hood and Peters (2004) have insulated three theoretical lenses for the analysis of paradoxes. The first is to look at unintended effects, i.e., the host of derivative and unpredicted problems that any deliberate social action generate (Merton 1936). The second way to look at paradoxes is the lens of cultural theory. Cultural theorists (Thompson et al. 1990), in fact, argue that what is seen as unanticipated or unintended is variable and socially construed. The third lens is to analyze paradoxes as system discontinuities and nonlinearities that stem out in the interaction among technology, human institutions, and social systems (Brooks 1986). This contribution adopts the lens of the unintended effects and focuses on the host of derivative and unpredicted problems that an outcome-based performance management system can generate (Maor 1999; Sieber 1981; Merton 1936).

3.2.2 Unintended Effects of Performance Management: Some Notable Theoretical and Empirical Contributions

The unintended effects of performance management are surely not a novelty in public management and administration literature. Many contributions, in fact, have classified and explored the various forms of unexpected developments and their determinants since early studies on administrative dysfunctions (Blau 1955). Providing a systematic review in the field is out of the scope of this chapter; however, some well-known contributions are mentioned here for the sake of a more complete argument, separating those that only provide analytical framework from those that add empirical evidence.

In the realm of taxonomies, an early example is the classification provided by Bouckaert and Balk (1991) who identify thirteen “diseases” of public productivity measurement. These diseases are the result of wrong assumptions underlying measurement, measurement errors, and problems concerning the content, position, and amount of measures. For example, authors debate whether it is therefore possible, desirable, or even necessary to measure public sector performance (Pangloss disease) because “government is efficient, because if it is not efficient, why hasn’t it already been changed?” (Bouckaert and Balk 1991). Furthermore, measuring public sector performance can disorient users and citizens. For example, “northern Great Britain seems to have more fires than other European countries because it has a better statistical technique for measuring” (Mandelbrot disease). Finally, authors call for managing the meaning of measurement, rather than focusing on measures themselves.

Another relevant and often cited classification is the one provided by Smith (1995) who classifies eight types of unexpected deviant behaviors. The first is “tunnel vision” which means choosing to concentrate on the easiest indicators and

ignoring the harder ones; “sub-optimization” means departments or units focusing on their performance incentives, disregarding the overall system’s performance; “myopia” consists in focusing on short-term targets at the expense of longer term objectives; “measure fixation” is focusing on the indicators and the metrics rather than the desired outcome; “misrepresentation” is either misreporting or distorting the data to create a good impression; “gaming” means deliberately under achieving in order to obtain a lower target next time; and finally, “ossification” occurs when no longer relevant indicators are not revised or removed. All these behaviors are, according to Smith, explained by four main factors, i.e., the divergence between the organizational objectives and the measurement scheme, the inability to accurately measure complex organizations, the inability to process performance data correctly, and the inability to respond to changing circumstances.

Van Thiel and Leeuw (2002) discuss how the performance paradox described by Meyer and Gupta (1994), a well-known problem in business firms, occurs in public sector organizations. Meyer and Gupta’s central idea is that there is a weak correlation between performance indicators and performance itself (Meyer and O’Shaughnessy 1993) since performance indicators run down over time. They, in fact, lose their value as measurements of performance and can no longer discriminate between good and bad performers. As a result, the relationship between actual and reported performance declines. Deterioration of performance indicators is caused by four processes (Meyer and Gupta, 1994, pp. 330–342), namely positive learning, perverse learning, selection, and suppression. Van Thiel and Leeuw (2002) claim that not only the paradox is recurrent in public sector organizations, but that the over-comprehensive mission of public sector organizations and the absence of a clear performance bottom-line are likely to reinforce the paradox (Fountain 2001; Torenvlied 2000; LeGrand 1991).

A parsimonious classification has been provided by Hood (2007a, 2006) who specifically uses the label “gaming” to define a family of strategic behaviors, aimed at maximizing positive feedbacks, regardless or at the cost of any other consideration. Three forms of gaming are conceptualized by Hood. The first is the ratchet effect, which is based on the expected tendency of target setters to set next year’s targets as an incremental advance over last year’s results. In this case, managers might have a perverse incentive not to exceed performance targets even if they could easily do so (Litwack 1993) or to negotiate relatively undemanding targets. The second type, threshold effects, refers to the effects of targets on the distribution of performance among a range of, and within, production units (Bird et al. 2005), putting pressure on those performing below the target level to do better, but also providing a perverse incentive for those doing better than the target to deteriorate to the standard. The last is output distortion, which refers to those behaviors aimed at “hitting the target (or winning the league) and missing the point” (Hood 2006). Output distortion happens when activities that are not measured are ignored or disregarded, when producers find ways to hit the targets or to improve their position in a ranking in ways that do not reflect the intentions of those who framed the system and finally when data reported to controllers are manipulated.

Beside the taxonomies, some contributions have performed empirical analyses, reaching contrasting results. Hood and colleagues (Bevan and Hood 2006; Hood 2006) have recurred to their classification to empirically analyze a specific setting, i.e., the British NHS. Their findings report recurring form of output distortions in the NHS, especially under those performance regimes based on target systems. On the other hand, using a similar empirical setting and a similar classification but different methods, Kelman and Friedman found no evidence of effort substitution (reducing effort on non-measured performance dimensions) or output distortion in their analysis of emergency room waiting times in the NHS. Also, the study by Wenger et al. (2008) has questioned the recurrent argument that there is an inevitable trade-off between quality and timeliness that leads to effort substitution or output distortion. In their analysis, based on US unemployment insurance system from 1997 to 2004, they found that the potential incentive for distortion was actually moderated by administrators who adopt management practices that facilitate improved outcomes for both timeliness and quality, leading to synergy between outcomes. Surely, inconsistencies in the debate call for further analysis in this field.

3.2.3 Under-Explored Facets: Rankings and Blame-Avoidance

Although nobody could reasonably claim that performance paradoxes are a novelty, literature did not fully address some angles. First of all, most previous contributions (with some exception, see Grizzle 2002) looked at the paradoxes induced by performance measures as such, while this chapter focuses solely on outcome measures. On the other hand, it might be argued that, since different types of measures create different accountability system, the kind of strategic response to them could vary as well.

Secondly, while there is a rich literature on the unintended effects of target systems, the unintended effects of rankings or intelligence systems are relatively less explored (Hood 2007a). This seems to be a relevant gap since rankings of public service performance have become a familiar part of the public management scene today and, like many relevant innovations, have their advocates and their detractors (Deming 2000; Gormley and Weimer 1999).

Thirdly, one potential cluster of unintended effects, namely blame-avoidance strategies by officials and service providers (Hood 2013; Weaver 1986, 1987), has been partly overlooked. Blame-avoidance is claimed to be central to both political and bureaucratic behavior (Hood 2013; Weaver 1986, 1987). Politicians and unelected officials are sometimes more motivated by avoiding blame for unpopular actions rather than taking credit for popular ones. Blame, in fact, is stickier than

credit, due to the negativity bias, i.e., the fact that people could be more impressed with losses, feared or real, than gains (Baumeister et al. 2001; Rozin and Royzman 2001; Lau 1985). Therefore, officials are likely to enact strategies to stay away from potential blame (Weaver 1986, 1987). Although there is no definitive account of the blame-avoidance strategies, Hood (2013) has classified three types of blame-avoidance approaches: presentational, agency, and policy/operational strategies. In the first case, decision makers try to “spin their way out of troubles” presenting issues (Druckmann 2011; Chong and Druckmann 2007) in a way that deflects blame. In the second case, they try to “find a scapegoat.” In other terms, officials design organizational architectures that shift responsibility for controversial matters to other players or make individual responsibility hard to detect. In the third case, they design decision-making procedures or operating routines that prevent them from making contestable judgements, minimizing the risk of incurring in blame.

Also in the case of blame-avoidance, beside taxonomies and conceptualizations, some empirical studies have been carried out. Many of them have traditionally been qualitative, particularly in the welfare-state literature (e.g., Pierson 1994; Lindbom 2007). There has been some quantitative analysis using survey data, notably on topics such as how political and institutional context affects blame of government for economic performance (Powell and Whitten 1993; Anderson 1995); how ministerial resignations absorb blame and raise government popularity (Dewan and Dowding 2005); observational and experimental studies of negativity bias (Dixon et al. 2013); and some other experimental work on the handling of blame (Sulitzeanu-Kenan 2006). Today, data available through digitally searchable media could make quantitative analyses of blame-avoidance and management more practicable than in the past (Hood et al. 2016), leading potentially to enhance generalizability.

While there is a vast literature on blame avoiding in fields such as communication strategy in crisis communication and crisis management (Boin et al. 2008, 2009, 2005), seldom (for a notable exception, see Charbonneau and Bellavance 2012 that looked at how politicians deal with blame from performance information) blame-avoidance has been conceived as a form of performance paradox and this is where this chapter adds a new focus.

What links the use of outcome measures to blame-avoidance, on the other side, is that the latter is somehow a mirror image of accountability, meaning that it occurs when actors fear to be blamed (Hood 2013). Without any accountability system in place, in fact, blame-avoidance would make no sense. The use of outcome measures is aimed, among other things, at making actors accountable as concerns the broader impacts of their policies (Van Dooren et al. 2010). This is why it might be argued that the shift to outcome measures is likely to somehow affect blame-avoidance behaviors by officials.

3.3 Methodology and Analytical Framework

3.3.1 Overall Research Strategy

The present contribution is based on the analysis of a pilot case in the Italian public sector, where a performance ranking composed of outcome measures was introduced as the pivotal performance management tool. The limitations of single case study as a research strategy are well known, especially as concerns the generalizability of results beyond the specific research site. However, some contributions have stressed that single case study can be considered an “ambitious inquiry” (Barzelay 1993) when the aim is to extrapolate “exemplar” and not just to produce locally valid explanations. The key conceptual resource that this approach uses is that of “social mechanism.” Social mechanisms can be defined (Hedstrom 2005; Hedstrom and Swedberg 1998) as unobserved analytical constructs that provide hypothetical links between observable events. In the words of the German political sociologist, Renate Mayntz (2004), mechanisms are sequences of causally linked events that occur repeatedly in reality if certain conditions are given. As Hedstrom and Swedberg (1998, p. 7) put it: “Assume that we have observed a systematic relationship between two entities, say I and O. In order to explain the relationship between them we search for a mechanism, M, which is such that on the occurrence of the cause or input, I, it generated the effect or outcome, O. The search for mechanisms means that we are not satisfied with merely establishing systematic covariation between variables or events [...]” In other terms, investigating the concatenation of social mechanisms means to open up the black box that leads from an event to another, unwrapping and dividing into smaller steps the cause–effect link that connects independent variable and outcome. This explains why the investigation of social mechanism has raised (somewhat creative) analogies with criminal investigation or trials. The concept of social mechanism, clearly grounded in sociology, has gained popularity in public administration and policy: A number of research programs (such as the process dynamics of public management policy-making) and research methodology paradigms (such as process tracing) heavily rely on the idea of discovering or testing social mechanisms. In this somewhat renewed light, in-depth single case studies can be a powerful tool to either generate hypotheses about the existence of a social mechanism or testing its functioning, moving beyond merely local explanations.

3.3.2 In Search of Mechanisms: Analytical Framework

In the attempt to insulate potential mechanisms that lead from the adoption of outcome measures to some form of blame-avoidance reaction, this chapter will employ the analytical lens provided by Beach and Pedersen (2013). In this light, a mechanism is made of “parts,” composed of “entities” that engage in “activities.”

	<i>Mechanism?</i>			
Ranking based on outcome measures <ul style="list-style-type: none"> • Naming and shaming • Enhancing transparency in allocation 	<i>Entity 1</i>	<i>Entity 2</i>	<i>Entity n</i>	Blame-avoidance strategies <ul style="list-style-type: none"> • Presentational strategies • Agencies strategies • Policy/operational strategies
	<i>Activity 1</i>	<i>Activity 2</i>	<i>Activity n</i>	




Fig. 3.1 Analytical framework

Entities are subject/object engaging in activities (noun), while activities are producers of change or what transmits causal forces through causal mechanisms (verbs). This analytical lens can be used both in an inductive way so to conceptualize potential concatenations of mechanisms from the field observation and in a theory testing fashion, so to test on the field the existence and functioning of potential mechanisms. The approach is used here in an inductive fashion: From the observation of the case, this contribution tries to insulate a mechanism to be field tested through future research protocols. Case analysis focuses, therefore, on a pilot case that suites analytically (Yin 2009) the research question. In the case under analysis, in fact, a ranking based on performance measures was adopted as the key performance tool and this provoked some blame-avoidance reactions. Therefore, the case allows to investigate the potential explananda, i.e., the mechanism that links the adoption of an outcome ranking (Arndt 2008; van de Walle 2008) to the kind of unintended reactions under analysis here (Fig. 3.1).

3.4 Exploratory Case Study

3.4.1 Context and Background

The analytical framework is used to observe a pilot case study in the Italian public sector at the intermediate tier of government (Region of Lombardy, northwest of Italy), where a performance ranking composed of outcome measures was adopted as the key management tool. Traditionally, Italy is classified as a Napoleonic country (Painter and Peters 2010; Peters 2008) in terms of administrative tradition. However, some typical features of this tradition have been undergoing significant changes. As concerns centralization, for example, devolution of powers toward the intermediate tiers has been significant over the years (Garlatti and Fedele 2014; Fedele and Ongaro 2008). On the other side, the institutionalization of managerialism varies across areas, although the traditional prevalence of administrative law vs management still holds true. Lombardy is considered, in many ways, as a

frontrunner of management logic within the Italian public sector and one of the regions that more radically took advantage of the room for maneuver opened by devolution reforms.

The policy under analysis is within the labor policy sector and stems from the integration, in 2013, of a number of separate policies in place before; its main objective is to place unemployed people back at work. The policy, like many others in Lombardy, is based on a vouchers arrangement, in which service provision is separated from its financing. The funding remains with government in the form of a voucher, which is issued to individual beneficiaries, entitling them to exchange the vouchers for the custom-tailored services provided by a range of public and private “employment services” suppliers. The providers need to be accredited in an open list composed and updated by the government so to ensure that providers comply to a set of quality standards. The institutions in charge for the policy are the “core” Department for Education, Training and Employment and the executive agency that operates at arm’s length, i.e., the Agency for Education, Training and Employment.

Although the policy had been, generally speaking, successful, placement results were worsened by gaming behaviors by service providers. Suppliers, in fact, were cream-skimming beneficiaries to attract the less problematic ones, i.e., those that cost less and are easier to place back at work. Some suppliers, furthermore, were locking in users just waiting to place them, without providing customized services.

3.4.2 A Partial Reform: Moving to a Performance Ranking

In 2013, the policy was changed along many lines. First of all, a number of previously existing policies were replaced by the current single multi-target policy. Two main innovations were introduced:

- individual users receive a score and are grouped in different “service streams” so to let the payment structure reflect differences across cases and avoid cream-skimming;
- every accredited supplier is assigned a “budget ceiling,” i.e., the maximum amount of public money that each supplier cannot exceed from the beginning of the policy to its end. The budget is dynamic inasmuch as periodically reviewed so to reduce locking-in and reduce market shares for underperforming providers.

The fact that typical gaming behaviors associated with a voucher environment were ruled out makes the case a “polar” one as concerns the observation of different clusters of unintended effects. For the policy round starting in October 2013, initial budget ceiling was allocated both on the basis of structural factors (65%) and of performance indexes (35%) as follows:

- 25% of the available funding was equally distributed among the operators;
- 40% of the available funding was attributed proportionally to the amount of public funding spent in the past rounds of the policy (as a proxy for the size of the operator);
- 28% of the available funding was attributed proportionally to the number of people put back at work;
- 7% of the available funding was attributed proportionally to the number of people placed at work only to the operators placing more than the average.

At the end of every bimonthly period, the budget is reviewed by the executive agency in charge for implementing the policy. Specifically, the agency simulates the reallocation of the unused budget on the basis of a pretty straightforward outcome measure, i.e., the number of people placed back at work. The redistribution of unused funding to the operators happens on the basis of three performance components:

- 60% of unused funding is distributed proportionally to the number of people put back at work;
- 20% of the unused funding is distributed proportionally to the number of people in the service stream 3 (most disadvantage) placed at work;
- 20% of unused funding is distributed proportionally to the number of relocated people only to the operators placing more than the average.

However, the executive agency in charge does not take a fully binding decision. It updates a composite performance ranking, based on the mentioned outcome measure, and sends it, along with a report containing specifics, to the parent department. The parent department faces a number of options: It can increase the budget ceilings when the unused budget is close to zero; can proceed with the allocation of the unused budget as per the simulation; or finally can do both things. A decision is then formally adopted by the parent department and made available to the public.

3.4.3 Unintended Effects: Blame-Avoidance Strategies

Case analysis allowed to document three blame-avoidance behaviors enacted by the different players involved. One significant strategy was put in place by service providers attempting at not losing positions in the ranking. Since the position in the ranking is associated with successfully treated cases, meaning persons that find a new job, any person not successful in finding a new job would worsen the providers' position in the ranking. This would worsen in turn their reputation among users and make them lose money tomorrow. A traditional gaming strategy, such as cream-skimming, was made impossible to pursue, due to the new systems of budget ceilings and service streams. "Pure" lock-in of single users was no longer convenient given the "dynamic" budget. Therefore, providers found a new form of

creative compliance: They would accept candidates, but would establish an unofficial waiting list. A candidate likely to be refused a job would be held back, until another, likely successful candidate, was ready to be spent for an interview along with the weaker candidate. Obviously, no official record was kept, leading to organizational amnesia that may be convenient for blame-avoidance in that no evidence of wrong-doing or dubious purposes could be found by inquiries or inspectors. This can be conceived as a special type of blame-avoidance behavior that incorporate elements of “output distortion.” The type of reaction described here is halfway between a traditional gaming strategy (output distortion) aiming at maximizing the incentives at the cost of any other considerations and a blame-avoiding behavior, aiming at preventing allegations of mismanagement and ineffectiveness by officials and users.

A second, and more clear-cut, case of blame game involved the executive agency and the parent department. Degrading a provider and reducing its budget is surely an unpopular decision. Delegation itself is an agency blame-avoidance strategy; therefore, the creation of an executive agency entails risk transfer from elected officials and parent department. However, doesn't matter how specific, every formal organizational design can leave room for interpretation and further routines to be agreed upon. Therefore, the executive agency insisted that its formal competencies should be limited to composing and updating the ranking and to producing budget simulations, while refusing to take on discretionary decision powers. The agency insisted that any discretionary power should instead be shifted to the parent department. The argument used by the executive agency's management was that the “formal” adoption of a decision like that is no longer an operational task but a political/regulatory one and therefore the parent department should be in charge. The argument resembles some administrative “mantra” about the policy/operational split (Fedele et al. 2007) in the department-agency relation. At the same time, this move might be interpreted as a blame-avoiding strategy of the agency type, i.e., carried out via organizational design. As a response, the parent department advocated, in the beginning, the compulsory application of the budgetary rules. In the arrangement advocated by the department, its competencies would be limited to formally adopt an act whose content was not the exercise of any discretionary power but the application of an automatic formula. This seems an attempt to enact protocolization, i.e., a blame-avoidance policy/operational strategy. Irrespectively to who won the blame game, however, the resulting arrangement spreads competencies among various organizations making detection of responsibility harder to trace in matters that can potentially attract blame from a relevant constituency, i.e., the providers.

The executive agency in charge for implementing the policy enacted another blame-avoidance strategy. The policy was pretty salient since unemployment is a crucial topic for public opinion in Italy. His dynamics are highly complex but the clear-cut metric adopted as the KPI was highly intelligible also for non-experts. One risk was that the policy could be deemed as uninfluential i.e., accused of placing a number of people back at work not exceeding the job market trends. This might lead to claims of wrong-doing and inefficiency by the media and the public opinion.

The agency, therefore, insisted on the introduction of a compulsory counterfactual impact evaluation to be carried out by external evaluators as a compulsory component of the policy cycle. Although this might be considered as a standard professional practice, a rival argument can be proposed. Since the policy had been designed intentionally as voluntary and not compulsory for unemployed individuals, self-selection was likely to ensure a positive feedback from a counterfactual, since only more qualified and motivated people would engage in the policy. Therefore, this strategy could be qualified as presentational, since it is based on data staging and framing although some element of agency strategy (i.e., delegation to third parties, i.e., the evaluators) are in place.

3.5 Analysis and Conclusions: Conceptualizing Mechanisms

Case analysis allows to provide some inputs useful to deepen the understanding of how outcome-based performance management influence blame-avoidance. The arguments presented here try to insulate a possible mechanism (Hedstrom 2005), whose sharpness, however, needs to be confirmed through further empirical research, especially through replication of case studies using the same analytical framework.

Juxtaposing the blame-avoidance strategies documented above, it can be noted that when the source of potential blame was users or the general public, the strategy enacted by both suppliers and various branches of government was of the presentational type, while, when the source of blame was service suppliers, blame-avoidance happened through agency and policy/operational strategies. Therefore, it could be argued that the nature of the group from which the blame could arise played a role in shaping the type of blame-avoiding reaction. In search for an explanation, it must be noted that a decision or a policy is not blame-worthy per se, but it is such in the eyes of a given group that find it unfair or damaging. This is why the nature of the “blamer” matters in terms of which strategy is chosen by a blame-avoiding player. In the case under analysis, it is evident that users and public opinion are, when compared to providers, a larger group that find harder to overcome the free-rider problem in activating collective action (Olson 1965); service providers are, instead, a special interest group that is likely to be more efficient in influencing the policy. Secondly—also on the basis of the first argument, users (not to mention general public) are more of an outsider to the policy village (Heclo and Widalwsky 1974) when compared to service providers; their participation and incorporation in the policy process does not seem very deep or institutionalized (Fung 2006). Finally, users and general public normally possess less specialist knowledge, again at least comparatively to administrators and providers. These features might account for explaining the type of blame-avoiding response. Presentational strategy in the end is about “telling and selling a story.” It might be argued that while it is viable to use this approach to convince a non-specialist-wide

Ranking based on outcome measures	Entity	Entity	Entity	Entity	blame-avoidance strategies
	Officials and service providers	Officials and service providers	Officials and service providers	Officials and service providers	
Naming and shaming	Activity	Activity	Activity	Activity	Presentational strategies
Enhancing transparency in allocation	Fears potential blame	Identify potential blaming groups	Weight blamers' knowledge and mobilization capacity	Estimate blamers' knowledge and mobilization capacity as not relevant	

Fig. 3.2 Mechanism a: from outcome measures to presentational strategies

Ranking based on outcome measures	Entity	Entity	Entity	Entity	blame-avoidance strategies
	Officials and service providers	Officials and service providers	Officials and service providers	Officials and service providers	
Naming and shaming	Activity	Activity	Activity	Activity	Agency strategies
Enhancing transparency in allocation	Fears potential blame	Identify potential blaming groups	Weight blamers' knowledge and mobilization capacity	Estimate blamers' knowledge and mobilization capacity as relevant	

Fig. 3.3 Mechanism b: from outcome measures to agency strategies

audience that possesses scant information and does not easily mobilize, an organized community of insiders is not likely to buy it and the strategy can backfire. Therefore, when dealing with organized insiders a rational blame avoider is likely prefers to recur to agency and policy strategies that are played beforehand through designing structures and procedures. Therefore, two potential mechanism, linking outcome measures (and their intended effects) to different blame-avoidance responses could be extrapolate from the analysis (Figs. 3.2 and 3.3)

Summing up, results contribute to the research agenda focusing on outcome-based performance management, exploring its relation with one cluster of unintended effects. More broadly, result can contribute to the research agenda exploring public management paradoxes. Secondly, on the practical side, results contribute to research on performance management in the public sector, providing insights on how to predict some traps in designing and using outcome measures.

However, it is necessary to point at findings' limitations that future research could address. First of all, since the present is a pilot case, meant at extrapolating a possible mechanism, the hypnotized concatenation of events and causal forces needs to be analyzed in a "mechanism-testing" fashion, through a larger number of empirical observation. In order to sharpen the findings, secondly, the mechanism could be tested across different administrative setting and policy areas. This could contribute, finally, to add knowledge on the specific topic analyzed here and finally to sharpen the paradigm of outcome-based performance management.

Appendix: Major Methodological Choices

Data have been collected in the first semester of 2016 through different steps, as described in the following. Multiple data collection techniques have been employed to allow triangulation and reach sharper evidence.

- Step 1. Analysis of official documents (aim: reconstructing the policy mechanism)

First of all, authors reviewed a number of administrative decisions shaping the functioning of the policy under analysis:

- D.g.r. n. 555/02.08.2013
- D.g.r. n. 748/04.10.2013
- d.d.u.o n. 9308/15.10.2013
- d.d.u.o. n.1436/24.02.2014
- d.d.u.o. n. 3591/29.04.14
- d.d.u.o. n. 3957/13.05.14
- d.d.u.o. n. 5186/17.06.14
- d.g.r. n.1983/20.06.14
- d.g.r. n. 2257/01.08.14
- d.d.u.o.n. 7587/05.08.14
- d.d.u.o. n. 11642/03.12. 2014
- d.d.u.o. n. 44/12.01. 2015
- Dgr n. 3144/18.02.2015
- d.d.u.o n. 1962/13.03. 2015
- d.d.u.o. n.2372/26.03. 2015
- d.d.u.o. n.3664/08.05.2015

Secondly, the research group reviewed the following documents:

- progress reports
- evaluation reports

Documents have been partly retrieved from the portal of Regione Lombardia and partly handed to the research group.

http://www.lavoro.regione.lombardia.it/cs/Satellite?c=Attivita&childpagename=DG_IFL%2FWrapperBandiLayout&cid=1213774160020&p=1213774160020&pagename=DG_IFLWrapper

- Step 2. In-depth interviews (aim: exploring the existence of potential mechanism)

After reconstructing the policy mechanism, the research group performed a first round of interviews with the director of the executive agency in charge for the implementation of the policy and his staff. The first round of interviews was not structured around a predefined set of questions so not allow potential interpretative bias.

- Step 3. Brainstorming (aim: insulating a potential mechanism)

After collecting field information, the research group have performed two rounds of brainstorming in order to insulate a potential mechanism from the empirical evidence, using the analytical framework described above (Beach and Pedersen 2013) and the procedures suggested by Miles et al. (2014) to handle qualitative data.

- Step 4. Semistructured interviews (aim: sharpening the potential mechanism)

After insulating potential mechanisms, the research group performed a second round of interviews with the director of the executive agency in charge for the implementation of the policy and his staff. The second round of interviews recurred to a semistructured set of questions aimed at sharpening the mechanisms insulated during step 3.

References

- Adcroft, A., & Willis, R. (2005). The (un)intended outcome of public sector performance measurement. *International Journal of Public Sector Management*, 18(5), 386–400.
- Anderson, C. (1995). *Blaming the government*. New York: M.E. Sharpe.
- Arndt, C. (2008). The politics of governance ratings. *International Public Management Journal*, 1(3), 275–297.
- Barzelay, M. (1993). The single case study as intellectually ambitious inquiry. *Journal of Public Administration Research and Theory*, 3(3), 305–318.
- Baumeister, R., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4), 323–370.
- Beach, D., & Pedersen, R. (2013). *Process-tracing methods: Foundations and guidelines*. Ann Arbor, MI: University of Michigan Press.
- Bevan, G., & Hood, C. (2006). What's measured is what matters: Targets and gaming in the English public health care system. *Public Administration*, 84(3), 517–538.
- Bird, S. M., Cox, D., & Farewell, V. T. (2005). Performance indicators: Good, bad, and ugly. *Journal of the Royal Statistical Society, Series A*, 168(1), 1–27.
- Blau, P. M. (1955). *The dynamics of bureaucracy*. Chicago: University of Chicago Press.
- Boin, A., 't Hart, P., & McConnell, A. (2009). Crisis exploitation: Political and policy impacts of framing contests. *Journal of European Public Policy*, 16(1), 81–106.
- Boin, A., 't Hart, P., Stern, E., & Sundelius, B. (2005). *The politics of crisis management: Public leadership under pressure*. Cambridge: Cambridge University Press.
- Boin, A., McConnell, A., & 't Hart, P. (2008). Governing after crisis. In A. Boin, A. McConnell, & P. 't Hart (Eds.), *Governing after crisis: The politics of investigation, accountability, and learning* (pp. 3–30). Cambridge: Cambridge University Press.
- Bouckaert, G., & Balk, W. (1991). Public productivity measurement: Diseases and cures. *Public Productivity & Management Review*, 15(2), 229–235.
- Bouckaert, G., & Halligan, J. (2008). *Managing performance: International comparisons*. London and New York: Routledge.
- Boyne, G. A., Meier, K. J., O'Toole, L., & Walker, R. M. (2007). *Public service performance: Perspectives on measurement and management*. Cambridge: Cambridge University Press.
- Brams, S. J. (1976). *Paradoxes in politics: An introduction to the nonobvious in political science*. New York: Free Press.

- Brooks, H. (1986). The typology of surprises in technology, institutions and development. In Clark W. C., & Munn, R. E. (Eds.), *Sustainable development of the biosphere* (pp. 325–350). Cambridge: Cambridge University Press.
- Charbonneau, E., & Bellavance, F. (2012). Blame avoidance in public reporting. *Public Performance & Management Review*, 35(3), 399–421.
- Chong, D., & Druckmann, J. N. (2007). Framing theory. *Annual Review of Political Science*, 10, 103–126.
- Deming, W. E. (2000). *The new economics for industry, government, education* (2nd ed.). Boston: MIT Press.
- Dewan, T., & Dowding, K. (2005). The corrective effect of ministerial resignations on government popularity. *American Journal of Political Science*, 49(1), 46–56.
- Dixon, R., Arndt, C., Mullers, M., Vakkuri, J., Engblom-Pelkkala, K., & Hood, C. (2013). A lever for improvement or a magnet for blame? Press and political responses to international educational rankings in four EU countries. *Public Administration*, 91(2), 484–505.
- Fedele, P., Galli, D., & Ongaro, E. (2007). Disaggregation, autonomy and re-regulation, contractualism: Public agencies in Italy (1992–2005). *Public Management Review*, 9(4), 557–585.
- Fedele, P., & Ongaro, E. (2008). A common trend, different houses: A comparative analysis of devolution in Italy, Spain and the United Kingdom. *Public Money & Management*, 2(2), 85–92.
- Fountain, J. E. (2001). Paradoxes of public sector customer service. *Governance*, 14(1), 55–73.
- Fung, A. (2006). Varieties of participation in complex governance. *Public Administration Review*, 66, 66–75.
- Garlatti, A., & Fedele, P. (2014). Autonomy, differentiation and performance: Lessons from a case study of civil service reform in P. In F. Visintin, D. Pittino, G. Lauto, & P. Mazzurana (Eds.), *Organizing for growth: Theories and practices* (pp. 95–101). Rome: BISAC: Business & Economics/Management.
- Gerrish, E. (2016). The Impact of Performance Management on Performance in Public Organizations: A Meta-analysis. *Public Administration Review*, 76(1), 48–66.
- Gormley, W. T., & Weimer, D. L. (1999). *Organizational report cards*. Cambridge, MA: Harvard University Press.
- Grizzle, G. A. (2002). Performance management and its dysfunctions. *Public Performance & Management Review*, 25(4), 363–369.
- Hecl, H., & Wildavsky, A. (1974). *The private government of public money*. Berkeley: University of California Press.
- Hedström, P. (2005). *Dissecting the social: On the principles of analytical sociology*. Cambridge: Cambridge University Press.
- Hedstrom, P., & Swedberg, R. (Eds.). (1998). *Social mechanisms: An analytical approach to social theory*. Cambridge: Cambridge University Press.
- Hood, C. (2006). Gaming in targetworld. *Public Administration Review*, 66(4), 515–522.
- Hood, C. (2007a). Public service management by numbers: Why does it vary? Where has it come from? What are the gaps and the puzzles? *Public Money and Management*, 27(2), 95–103.
- Hood, C. (2007b). What happens when transparency meets blame-avoidance? *Public Management Review*, 9(2), 191–210.
- Hood, C. (2013). *The blame game: Spin, bureaucracy, and self-preservation in government*. Princeton, NJ: Princeton University Press.
- Hood, C., Jennings, W., & Copeland, P. (2016). Blame avoidance in comparative perspective: Reactivity, staged retreat and efficacy. *Public Administration*, 94(2), 542–562.
- Hood, C., & Peters, B. G. (2004). The middle aging of new public management: Into the age of paradox? *Journal of Public Administration Research and Theory*, 14(3), 267–282.
- Kelman, S., & John, N. (2009). Performance improvement and performance dysfunction: An empirical examination of distortionary impacts of the emergency room wait-time target in the English National Health Service. *Journal of Public Administration Research and Theory*, 19(4), 917–946.

- Lau, R. R. (1985). Two explanations for negativity effects in political behaviour. *American Journal of Political Science*, 29(1), 119–138.
- LeGrand, J. (1991). The theory of government failure. *British Journal of Political Science*, 21(4), 423–442.
- Lindbom, A. (2007). Obfuscating retrenchment: Swedish welfare policy in the 1990s. *Journal of Public Policy*, 27(2), 129–150.
- Litwack, J. M. (1993). Coordination, incentives and the Ratchet effect. *The Bell Journal of Economics*, 24(2), 271–285.
- Maor, M. (1999). The paradox of managerialism. *Public Administration Review*, 59(1), 5–18.
- Massaro, M., Dumay, J., & Garlatti, A. (2015). Public sector knowledge management: A structured literature review. *Journal of Knowledge Management*, 19(3), 530–558.
- Mayntz, R. (2004). Mechanisms in the analysis of social macro-phenomena. *Philosophy of the Social Sciences*, 34(2), 237–259.
- Merton, R. K. (1936). The unanticipated consequences of purposive social action. *American Sociological Review*, 1(6), 894–904.
- Meyer, M. W., & Gupta, V. (1994). The performance paradox. *Research in Organizational Behavior*, 16, 309–369.
- Meyer, M. W., & O’Shaughnessy, K. (1993). Organizational design and the performance paradox. In R. Swedberg (Ed.), *Explorations in economic sociology* (pp. 249–278). New York: Russell Sage Foundation.
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Thousand Oaks, CA: SAGE Publications Inc.
- Nielsen, P. A., & Moynihan, D. P. (2016). How do politicians attribute bureaucratic responsibility for performance? Negativity bias and interest group advocacy. *Journal of Public Administration Research and Theory*, 27(2), 269–283.
- Olson, M. (1965). *The logic of collective action: Public goods and the theory of groups*. Cambridge, MA: Harvard University Press.
- Painter, M. J., & Peters, B. G. (2010). *Tradition and public administration*. London: Palgrave Macmillan.
- Peters, B. G. (2008). The napoleonic tradition. *International Journal of Public Sector Management*, 21(2), 118–132.
- Pierson, P. (1994). *Dismantling the welfare state? Reagan, thatcher and the politics of retrenchment*. Cambridge: Cambridge University Press.
- Powell, G. B., Jr., & Whitten, G. D. (1993). A Cross-national analysis of economic voting: Taking account of the political context. *American Journal of Political Science*, 37(2), 391–414.
- Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance and contagion. *Personality and Social Psychology Review*, 5(4), 296–320.
- Sieber, S. (1981). *Fatal remedies*. New York: Plenum.
- Smith, P. (1995). On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration*, 18, 277–310.
- Sulitzeanu-Kenan, R. (2006). If they get it right: An experimental test of the effects of the appointment and reports of UK public inquiries. *Public Administration*, 84(3), 623–653.
- Thompson, M., Ellis, R., & Wildavsky, A. (1990). *Cultural theory*. Boulder, CO: Westview.
- Torenvlied, R. (2000). *Political decisions and agency performance*. Dordrecht: Kluwer Academic Publishers.
- Van de Walle, S. (2008). What services are public? What aspects of performance are to be ranked? The case of “Services of General Interest”. *International Public Management Journal*, 11(3), 256–274.
- Van Dooren, W., Bouckaert, G. & Halligan, J. (2010). *Performance management in the public sector*. London, New York: Routledge.
- Van Thiel, S., & Leeuw, F. (2002). The performance paradox in the public sector. *Public Performance & Management Review*, 25(3), 267–281.
- Weaver, R. K. (1986). The politics of blame avoidance. *Journal of Public Policy*, 6(4), 371–398.
- Weaver, R. K. (1987). The politics of blame. *The Brookings Review*, 5(2), 43–47.

- Wenger, J. B., O'Toole, L. J., & Meier, K. J. (2008). Trading speed for accuracy? Managing goal conflict and accommodation in the U.S. unemployment insurance program. *Policy Studies Journal*, 36(2), 175–198.
- Yin, R. K. (Ed.). (2009). *Case Study Research. Design and Methods* (4th ed.). Thousands Oaks: Sage.

Chapter 4

Designing Outcome-Based Performance Management Systems to Assess Policies Impacting on Caesarean Section Rate: An Analysis of the Sicilian Maternity Pathway

Enzo Bivona and Federico Cosenz

Abstract The reduction in Caesarean sections (CSs) is widely considered a priority in the public decision makers agenda. Though the World Health Organisation has strongly encouraged countries to implement policies to reduce CSs to 10–15%, after almost thirty years this goal appears still far from its achievement. The literature depicts CS as a multifaceted phenomenon whose causes involve different factors, ranging from the patient sphere to the health care level of services provided, and the societal preference of CS practice. Policy makers aiming to standardise cares and to reduce CSs often implement maternity pathways (MP). By investigating the MP introduced in the Sicilian region, the authors highlight the need to adopt an outcome-based performance management approach to assess the effectiveness of CS reduction policies. The suggested perspective also reveals the necessity to frame and coordinate the interdependencies between the different actors playing a crucial role in the MP.

Keywords Maternity pathway · Caesarean sections · Dynamic performance management · Public policy analysis · Outcome measures

E. Bivona (✉) · F. Cosenz
Department of Political Sciences and International Relations, CED4 System Dynamics Group, University of Palermo, Palermo, Italy
e-mail: enzo.bivona@unipa.it

F. Cosenz
e-mail: federico.cosenz@unipa.it

4.1 Introduction

High Caesarean birth rates are still an international concern. In 2011, 1 of 3 women who gave birth in the USA did so by Caesarean delivery. In the same period, European countries recorded on average a lower rate, ranging from a minimum of 14.7% in the Republic of Iceland up to 37.7% registered average in Italy. Though recent data show a slight reduction, Caesarean delivery rates are far from the 10–15% considered the ideal rate by the World Health Organisation (WHO 1985). The appropriate Caesarean Section (CS) rate is still a hot topic in debates among professionals on one side and public policy makers on the other side. For certain critical clinical conditions (such as placenta previa or uterine rupture), Caesarean delivery is firmly established as the safest delivery. However, for most low-risk pregnancies, Caesarean delivery appears to pose greater risk of maternal morbidity and mortality than vaginal delivery. It also implies higher costs for both health care providers and citizens.

The wide range of CSs case-mix makes the interpretation of results a very hard task. It is clear indeed that CS rates should not be judged in isolation from other characteristics, such as the epidemiological profiles of the patients and the pre- and post-partum health care services provided.

To reduce CS rates, a multidisciplinary approach is widely advocated. Previous research often focused on practices professionals must adopt to reduce CS rates. Other studies investigated how cultural norms in a given community are likely to impact on the acceptance of CSs.

This study proposes a different perspective of analysis. In particular, it aims to investigate the critical role played by the maternity pathway adopted by the Sicilian region to reduce CS rates.

Based on this analysis, we propose the adoption of a performance management approach to design a set of outcome measures aimed at supporting decision makers in assessing the impact of CS reduction policies.

Due to the complexity of the CSs, an approach aimed to identify the drivers favouring or discouraging the recurrent use of this practice is indeed required. To this aim, a dynamic performance management approach is outlined, and the relationships between strategic resources accumulation and depletion processes, performance drivers and end results impacting on CSs are made explicit.

The chapter is divided into five sections. We begin with an analysis of the existing literature on the main drivers affecting CSs. In the next section, we analyse the maternity pathway recently introduced in Sicily. Based on the limitations showed by the maternity pathway measures, the dynamic performance management approach oriented to design a set of outcome measures impacting on the CS rate is illustrated. Then, such an approach has been applied to assess the policies impacting on CS in the Sicilian context. In the final section of the chapter, conclusions and research perspectives are highlighted.

4.2 What Drivers Affect Caesarean Rate?

The WHO (1985) set a justifiable CS rate in any region of the world not greater than 10–15%. Some countries, especially in Northern Europe, such as Finland, Norway and the Netherlands record an average Caesarean rate around 17%, and they may reasonably reach the WHO target. However, they can be considered an exception. USA, Asia and many other European countries cope with a higher CS rate, ranging from 20 up to 35% (Macfarlane et al. 2016). For these countries, the WHO's CS goal may appear as a utopia.

Several studies investigated those factors underlying the raise worldwide in the CS rate. By analysing such contributions, it is possible to find a general consensus around the main causes determining the growth in the CS rate. However, due to the multifaceted complexity of the investigated phenomenon, it is hard to detect the same level of agreement on policies able to reduce it. CS causes are various: they range from pregnant/unborn health conditions, health care practices, cultural propensity towards CS, legal issues up to financial incentives, just to mention few of them (Brennan et al. 2009; Zizza et al. 2011; Betrán et al. 2007; Li et al. 2013, 2017).

A study investigating extensively such causes is far from the goal of this research, and it might run the serious risk to be incomplete. This work, instead, tries to offer a critical review of those factors that can be targeted by decision makers to implement successful strategies to reduce the use of CSs.

Looking at CS causes potentially addressed by decision makers, a recent study highlighted the role played by the *clinical dimension* (i.e., the way pregnant women receive hospital cares), in terms of diversified obstetrical practices diffused inside hospitals and poor university trainings offered to professionals on the practical aspect (Istituto Superiore della Sanità 2010).

Other studies remarked the influence that *social/cultural dimension* has on the woman decision to have a CS. The social/cultural misperception that the CS is often safer than the vaginal delivery (VD) and the declining women birth rate, due to a decrease in women fertility recorded worldwide, has been indicated as main factors stimulating CS practice. This scenario may influence obstetricians, who might be conditioned by non-medical risk factors on the decision making process (Triunfo et al. 2015).

Robson et al. (2013) recently stressed the importance to take into account the *patient dimension*. Although it is generally accepted that the health conditions of both the pregnant and the unborn have a critical role in the decision to make a CS or to favour a VD, the increasing autonomy of the women to decide the mode of delivery may significantly impact on the request of a CS. In most national guidelines, we recognised the reinforced right of women to decide for a Caesarean delivery even in cases of low-risk pregnancies and in presence of obstetrician disagreement.

The brief review of the above studies depicts a multidimensional picture of those factors impacting the CS rate. Designing effective strategies to reduce the CS rate is

a very complex task. This difficulty is exacerbated at the level of individual health facility, as it is often difficult to determine an appropriate rate of CS. In fact, differences in the groups of patient's case-mix and related obstetric profile make inapplicable a universal reference rate for CS, and consequently, decision makers may find difficult to interpret results and to design unique effective policies to reduce CS rates.

4.3 An Analysis of the Maternity Pathway in the Sicilian Region

Sicily is a region located in Southern Italy of about 5 millions of inhabitants. In 2015, the regional government adopted the national guidelines to implement in the health care system a maternity pathway (MP). It aims to coordinate the maternity care services provided by the different perinatal players operating in a given province. It is conceived as a care pathway aimed at fostering the integration between hospitals and other health care players operating within the same local areas.

This pathway begins at the pre-conception phase—including pregnancy and childbirth—and it ends in the early months of child's life. It is designed to provide a standardised procedure to pregnant women with the intent to respect the regulation on woman and child health with a focus on reducing CSs. Its articulation depends on the level of pregnancy risk¹ and, based on this, distinguishes alternative solutions. In particular, neonatal risks are classified in: (1) maternal, (2) foetal and (3) maternal/foetal. The pregnancy risk level is determined according to the WHO model (WHO 2001).

Following a systemic perspective between both maternal cares and psychosocial aspects, the pathway encompasses a local network (birth pathway network) made of regional and hospital services connected to the mother-child area. In particular, this network includes: family counselling network, natal centres (divided in first and second level of pregnancy risk), general medicine doctor network (MMG), free-choice paediatricians (PLS), neonatal emergency transport service (STEN), maternal care emergency transport service (STAM), regional offices located inside hospital units. Each natal centre is divided into two areas: obstetrical and neonatal/paediatric. Second-level natal centre is also known as neonatal intensive care unit (UTIN).

¹Examples of first-level pregnancy risk are: <16 or 40> years old woman; <150 cm height; cervical or vaginal infections; foetal malformation; three or more spontaneous abortions; epilepsy. On the other hand, examples of second-level pregnancy risk include: vaginal bleeding, uterine malformations, uterine myomas and presence of pelvic mass. There are also third-level pregnancy risks—directly treated in natal centres—such as twin pregnancy, foetal growth delays, foetal malformations and diabetes.

The main steps of the MP are:

1. pre-conception period organisation and management;
2. preliminary pregnancy risk assessment (obstetrical and psychosocial);
3. low-risk pregnancy cares;
4. higher risk pregnancy cares according to the related specific risk conditions;
5. training courses leading to childbirth;
6. cares to pregnant women affected by psychological or social difficulties;
7. postpartum cares (perineum assessment, breastfeeding guidance and support, contraceptive advisory, postpartum depression early diagnosis);
8. neonatal cares.

A graphical representation of the MP adopted in the Sicilian region is portrayed in Fig. 4.1.

Particularly, pregnant women may choose to receive pregnancy cares by private gynaecologists or by family counselling, which preliminarily filters physiological and pathological pregnancies. Focusing on the latter case, two pathways are recommended according to the risk analysis:

- *Physiological pregnancy*: in this case the pregnant woman undergoes a series of clinical exams in the family counselling and, after 36 weeks and 6 days (i.e., the end of pregnancy), goes to a first-level natal centre located inside a hospital where to give birth. Afterwards, the mother goes back to the family counselling to undergo treatments (such as breastfeeding, contraceptive and postpartum phases) oriented to support the postnatal progress.
- *Pathological pregnancy*: in this case the pregnant woman—who shows pathologies or risk factors before the 34th week—is entrusted to UTIN which provides support to premature births and specialised cares to specific pathologies. Namely, in case of low pregnancy risk, maternal cares to the pregnant woman are provided by family counselling until the 36th week; then, childbirth

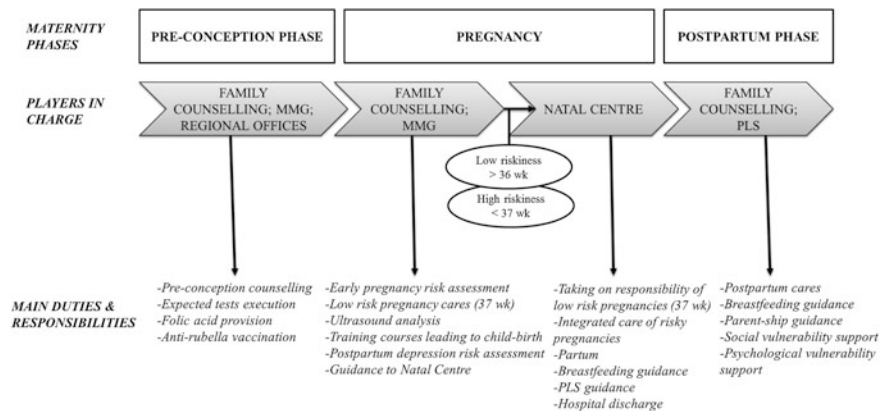


Fig. 4.1 The MP adopted in the Sicilian region

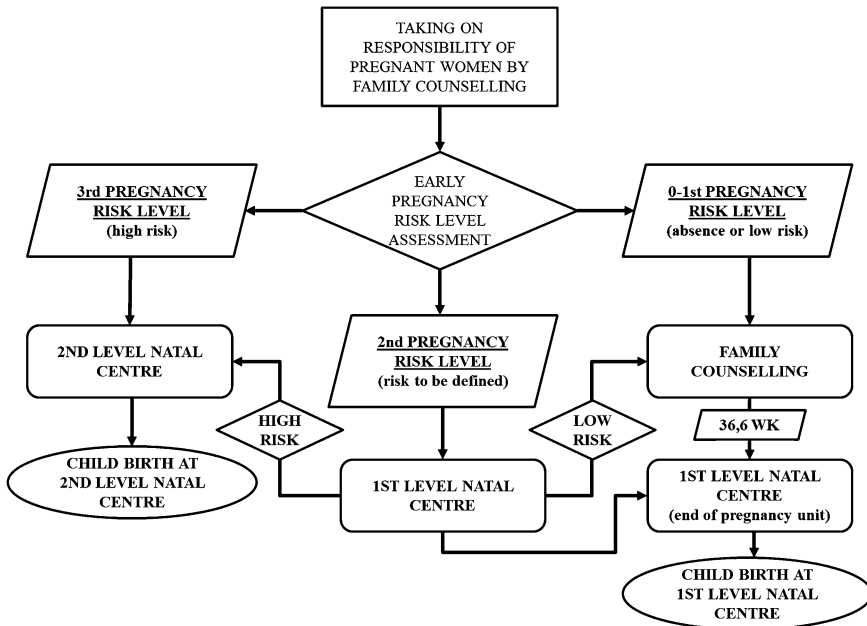


Fig. 4.2 The MP map based on risk level classification

is carried out in first-level natal centres. On the other hand, in case of high pregnancy risk, the pregnant woman is directly entrusted to second-level natal centres where, at the end of pregnancy, the baby will born. Postnatal services are performed by the family counselling.

Figure 4.2 synthesises the MP map based on the above risk level classification.

In addition, all players involved in the MP have to implement specific performance measurement systems to assess the different activities carried out.

However, the indicators included in these performance measurement systems often show a number of limitations mainly due to a bureaucratic perspective used in their design. In particular, they aim at capturing the increase in the use of MP services by pregnant women, but this does not ensure a reduction in the CS rate per se. It might represent a useful measurement once the validity of such a pathway in reducing CS rate is fully demonstrated. In addition, they do not support decision makers to understand how to change the resources allocation policies to affect the CS rate.

To effectively support decision makers' understanding of the outcomes resulting from the implementation of a given policy, a measurement system should identify and measure those causal determinants producing a major effect on CS rate. Namely, according to a systemic perspective such drivers are strictly related to resource allocation policies and are likely to affect performance, in terms of both

outputs and outcomes over time. Therefore, monitoring performance drivers may serve as a basis on which to promptly reformulate public policies.

In the next section, we introduce a method to design more effective outcome measures to support decision makers in assessing the impact of regional policies on CS rate.

4.4 A Dynamic Performance Management Approach to Design a Set of Outcome Measures Impacting on CS

Bringing outcome measures into public service decision making and management is nowadays a key challenge to take on a broader perspective of public performance management results, as well as to ensure increasing benefits to the territorial area in terms of quality of life and welfare (Pollitt and Bouckaert 2004). The importance of measuring outcomes in the public sector relies on the fact that, unlike the private sector, there is no bottom line against which performance can be measured. In fact, while assessing short-term results of a single institution is generally considered feasible (output), problems occur when we aim to measure the long-term impact produced by the aggregated contribution—in terms of output—of many public/private organisations on the local area in which they operate (outcome). As Bianchi et al. (2017) assert “the use of a short-term perspective and a sectoral approach in the formulation and implementation of strategies...lead to a static view of the system and to a lack of coordination in policy making between different public agencies, non-profit and private stakeholders.”

This approach is unlikely to help policy makers to identify sustainable actions on complex issues which span across several jurisdictions, both in terms of level (e.g., national, regional, local) and policy domain (e.g., policing, welfare, education, justice). Indeed, the complex interaction between these players, an idiosyncratic perspective of public performance management, and the lack of a “robust” coordination, generate critical methodological issues to design and implement outcome-based performance measurement systems.

The health care sector has a long tradition in using outcome measures to assess care quality. Such a tradition is mainly rooted in analysing performance according to a medical approach.² On this concern, several perinatal outcome indicators have also been designed and implemented, particularly in Europe (Wildman et al. 2003). From a strategic management perspective, to overcome the above constraints

²In medical science, the term “outcome” has to be intended in the Donabedian’s conceptualisation, i.e., it refers to a patient’s health status or change in health status resulting from the medical care received (Donabedian 2005). This definition is oriented to analyse the post-cares patient survival conditions and includes intended outcomes (e.g., the relief of pain), as well as unintended outcomes (e.g., complications).

to outcome measurement design, coordination between different services and organisations (internal and external coordination) and a methodological approach to performance management are required. As Fixsen et al. (2005) also remark, this coordination helps to capture the multi-level influences on performance measurement implementation, from external influencers to organisational and core implementation process components. The critical role played by implementation theories has been largely debated in the health care management literature (Damschroder et al. 2009).

Implementation is a social process directly connected with the context in which the adoption and use of a new practice take place, where the context includes a set of circumstances or factors that affect implementation. It represents the gateway between an organisational decision to adopt an intervention—such as the introduction of outcome-based performance management systems applied to MP—and the routine use of that intervention (Dixon-Woods et al. 2011). As such, it focuses on the transition period during which targeted stakeholders become increasingly skilful, consistent and committed in their use of an intervention (Klein and Sorra 1996; Damschroder et al. 2009). Implementation may vary according to a number of factors, such as the characteristics of the intervention, the governance levels involved and the organisational setting.

Unlike other health care interventions (Greenhalgh et al. 2004; Damschroder et al. 2009), the implementation of MP requires an active change process aimed at fostering its use at individual, organisational and inter-institutional level. It also implies the adoption of a performance measurement system calibrated on the multiple sub-processes encompassing the MP, and related interplays, leading to outcomes.

A dynamic and outcome-based performance management approach is particularly valuable for this purpose, since time disjunctions between actions and results, and nonlinear feedback relationships affecting outcomes, limit decision makers' understanding of the structure and behaviour of the system in which their policies will be implemented (Bianchi et al. 2017). This approach supports them to manage possible risks related to unintended effects of policies which, although they may look consistent from a static and sectorial perspective, may fail in the long term due to a lack of coordination or lack of flexibility (Bianchi et al. 2017; Ghaffarzadegan et al. 2011).

A dynamic and outcome-based performance management approach is adopted to support decision making through better coordination between performance measurement reporting and strategy/policy design. Such coordination helps policy makers and public managers to trace both causes and drivers that have led to a given performance level over time. It also contributes in enhancing the diagnosis process to put in place corrective actions and strategies oriented to fill the gap between the actual and the target performance. As such, similarities between this approach and the Balanced Scorecards by Kaplan and Norton (1992) applied to health care can be found especially in terms of purposes and contents (Zelman et al. 2003). In fact, both of them are grounded on the assumption by Nelson et al. (2002, p. 18) according to which “change in the health system is subject to a linked chain

of effect that connects individual patients, communities and clinicians with small, naturally occurring front-line units, with countless large and small host organisations all of which exists in a modulating policy, legal, social, financial and regulatory environment.” In addition, they both are concerned with the challenge of implementing processes—e.g., clinical pathways—consistently across a large but extremely diverse organisation—e.g., health care sector institutions (Bilkhu-Thompson 2003).

In particular, an outcome-based performance management approach is primarily concerned with the identification of both end results (output and outcome) and their respective drivers. To affect such drivers, public organisations must build up, preserve and deploy a proper endowment of strategic resources that are linked each other. This also implies that decisions made by different actors upon interdependent strategic resources should be coordinated each other according to a systemic view. Particularly, each strategic resource should provide the basis to sustain and foster others in the same system. For instance, both physicians and technological equipment provide cares, which affect perceived service quality. This produces a certain (i.e., positive or negative) impact on the hospital reputation which, in turn, influences patient satisfaction. A change in patient satisfaction will affect public funding to increase the stock of available financial resources, and eventually care and service quality (see Fig. 4.3).

Performance drivers are associated with critical success factors in the referring public sector. They can be measured in relative terms—as a ratio between the

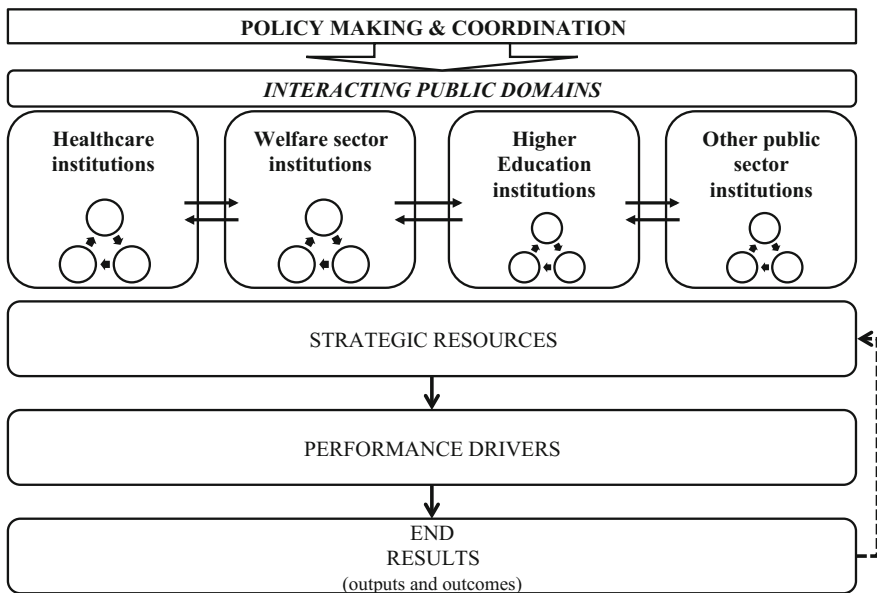


Fig. 4.3 Outcome-based Dynamic Performance Management framework (adapted from Bianchi 2016, p. 73)

organisational performance perceived by users and a benchmark—or a target value. Such a denominator must be gauged in relation to perceived past performances or users' expectations.

Following this approach, it is possible outlining the policy options formulated to affect the strategic resources that will influence performance drivers, and—through them—the end results, which in turn will feedback on the strategic resources. In addition, this performance management perspective does not limit its relevant boundaries to a single organisation. Rather, it is aimed at designing performance measures that can assess the long term effect and broader impact of implemented policies by a single player on a much wider system. A system-wide view of performance eventually requires to be combined with an internal view, by each organisation, in order to foster a strategic dialogue and coordination among the key players oriented to improve their aggregated contribution to the overall system.

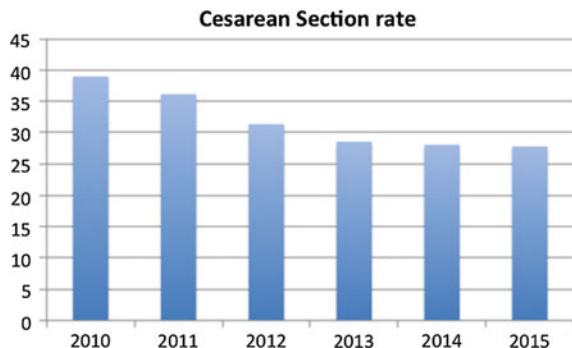
4.5 Applying the DPM Approach to Design a Set of Outcome Measures to Assess Policies Impacting on CS in the Sicilian Context

In the last decade in Italy, all regional governments have been required to set the CS rate around 20%, in particular for those women who are pregnant for the first time (*primiparous*). Since 2010, this rate has been slightly decreasing from 28.3 to 25.7%. However, there are several regions in which such a rate reaches values above the 50%.

In Sicily, in 2010 the recorded CS rate was almost 40%. Nowadays, it is stable around 28%, still far from the 20% goal set at national level (Fig. 4.4).

In 2015, the regional government introduced a MP to standardise the care and to reduce the CS rate. With the introduction of such a pathway, some indicators have been identified. These measures track the diffusion of the pathway used by pregnant women, but they are not able to assess the effectiveness of the resource allocation

Fig. 4.4 CS rate recorded in Sicily in the period 2010–2015. *Source* Regional Health care Department, Sicily



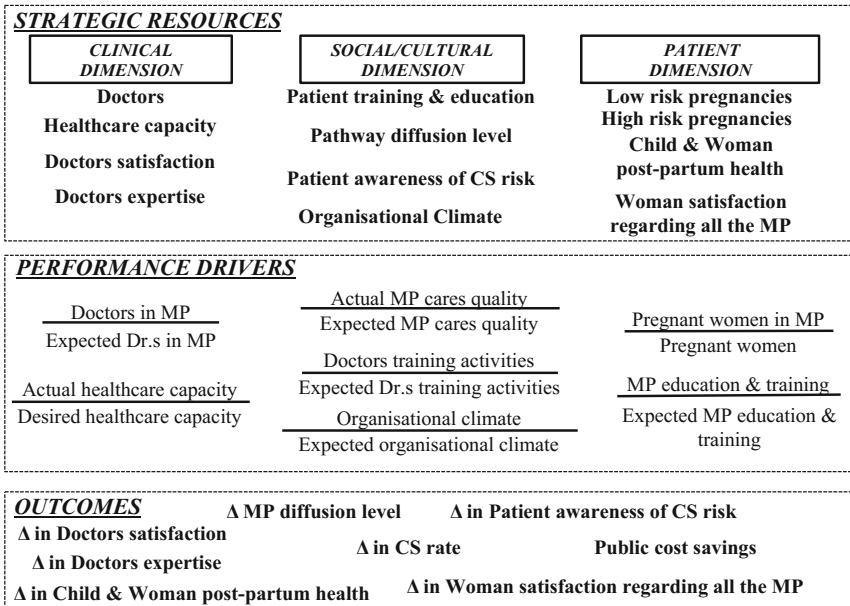


Fig. 4.5 Outcome-based Dynamic Performance Management applied to MP

policies oriented to reduce the CS rate. With the intent to overcome such shortcomings, an outcome-based performance management approach is here used.

The framework reported in Fig. 4.5 outlines the relationships between outcomes, performance drivers and strategic resources affecting CS rate.

Among the outcomes, different measures have been identified according to the *clinical*, the *social/cultural* and the *patient* dimension impacting on a change in CS rate, as emerged from the literature discussed in Sect. 4.2. In particular, in the clinical dimension, two main outcomes are taken into account: change in doctors’ satisfaction and in doctors’ expertise. They depend on the decision to invest in doctors’ training, as well as in organisational climate.

In the social/cultural dimension, change in patient awareness of CS risk and change in MP diffusion level aim to detect the effectiveness of the promotion and broadcasting actions undertaken by decision makers to impact on CS rate. The patient dimension includes change in child and woman post-partum health, and change in women satisfaction regarding all the MP phases, as a consequence of the initiatives aiming to improve the quality and the number of services of the MP. All the above measures are likely to affect the main outcome indicator resulting in a change in CS rate, which in turn influences public cost savings.

To understand how policies undertaken by decision makers are able to affect the above outcomes, it is worth making explicit the underlying performance drivers. Figure 4.5 shows a number of performance drivers. For instance, the ratio between actual and expected doctors’ training activities is likely to capture a change in

doctors' expertise. In addition, the alignment of MP education and training initiatives with the expected level drives a change in the outcome patient awareness of CS risk. Likewise, the ratios between actual and total MP care, on the one side, and between actual and expected MP cares quality, on the other, produce an effect on a change in women satisfaction.

The above performance drivers indeed cannot be directly influenced by decision makers. In fact, such measures result from a change in the interrelated strategic resources. Strategic resources vary through accumulation and depletion processes, which can be either affected by decision makers and organisational routines. For instance, the number of doctors can be increased by the implementation of a hiring policy, while doctors' satisfaction changes as a consequence of internal routines related to the organisational climate. Among the other strategic resources, there are patient training and education, pathway diffusion level, patient awareness of CS risk, low- and high-risk pregnancies, woman satisfaction regarding all the MP phases and organisational climate.

Figure 4.5 also shows a feedback relationship between the different elements of the DPM framework. In particular, a change in the outcome measures is likely to influence the corresponding strategic resources, which in turn affect the associated performance drivers giving rise to a cause-and-effect chain (e.g., change in women satisfaction regarding all the MP → woman satisfaction regarding all the MP → ratio “pregnant women in MP/pregnant women” → change in CS rate change in woman satisfaction regarding all the MP). Making explicit the above feedback interdependences provides decision makers with a deeper understanding of the main mechanisms driving the outcomes, and consequently, policies can be reformulated by taking into account such a perspective.

4.5.1 Implications Arising from the DPM Implementation in Assessing CS Reduction Policies

The DPM framework previously described highlights the need to coordinate the multiple actors (regional offices, family counselling and natal centres) intervening in the MP to design and implement effective and sustainable policies aimed at reducing the CS rate. It also helps policy makers to identify the key factors they can alter, i.e., the strategic resources reported on the top of the DPM framework.

To reduce the CS rate, regional offices may invest in promoting the “pathway diffusion level.” As a consequence, a high number of pregnant women enter the MP and, after an assessment of the patient low or high risk profile, the pregnant woman is assigned to a corresponding MP. However, to perform such a timely patient profile diagnosis, the family counselling carrying capacity (e.g., doctors and health care capacity) should match the demand increase of MP care. A lack of such a strategic resource negatively affects the service care level expected by pregnant women—captured in the DPM through the driver “actual MP cares

quality/expected MP cares quality”—which in turn impacts on the outcome “change in woman satisfaction regarding all the MP.” This will feed back on the strategic resource “pathway diffusion level,” thereby decreasing the number of pregnant women in the MP. In the medium–long term, if investments in family counselling, doctors and health care capacity are neglected, efforts by regional offices in promoting the pathway diffusion will be unable to generate the desired outcome.

Family counselling, in addition to the investments in capacity, can also allocate resources to train and educate pregnant women to foster their awareness about CS-associated risks and VD potential benefits, which may significantly reduce women request to have a CS in favour of a VD.

The efforts made in promoting MP diffusion, patient’s risk profile diagnosis and education, made by regional offices and family counselling respectively, may be insufficient to achieve the desired outcome, if the level or type of strategic resources that natal centres have (such as doctors, doctors expertise, doctors satisfaction and health care capacity) are not aligned to the demand of MP care. In fact, investing in doctors’ recruiting and training, beds and equipment, and organisation procedures may affect both the patients, in terms of quality of care received, and the doctors’ expertise and satisfaction. An increase in the performance drivers “doctors in MP/expected doctors in MP” and “doctors training activities/expected doctors training activities” produces a positive impact on the outcomes doctors’ expertise and satisfaction, leading to an improvement in the CS rate.

4.6 Conclusions

This chapter outlined a critical analysis of the MP adoption in the Sicilian region. The authors propose to adopt the DPM approach to identify a set of outcomes able to support decision makers in designing policies to reduce the CS rate.

The literature analysis depicts a very complex picture of those factors influencing the change in CS rate. Such causes can be summarised into three main dimensions: clinical, social/cultural and patient. This classification has allowed the identification of interrelated policies decision makers can implement to act on strategic resources. Such resources are likely to generate a change in performance drivers, which in turn affect outcomes. We discussed a number of outcomes impacting on CS rate, and they can be used to gauge the effectiveness of MP implementation over time.

One of the main limitations of this chapter is that it explores a conceptual articulation of the MP, as it has been designed by the regional administration. To validate the suggested framework, we will aim to provide an empirical evidence through a field investigation in one Sicilian province in which the MP has been already implemented.

References

- Betrán, A. P., Meriáldi, M., Lauer, J. A., Bing-Shun, W., Thomas, J., Van Look, P., et al. (2007). Rates of cesarean section: Analysis of global, regional and national estimates. *Paediatric and Perinatal Epidemiology*, *21*, 98–113.
- Bianchi, C. (2016). *Dynamic Performance Management*. Zurich, Switzerland: Springer International Publishing.
- Bianchi, C., Bovaird, T., Loeffler, E. (2017). Applying a dynamic performance management framework to wicked issues: How co-production helps to transform young people's services in Surrey County Council, UK. In E. Borgonovi, W. Rivenbark, C. Bianchi, (Eds.), Exploring the Complexities of Performance Management (forthcoming), *International Journal of Public Administration*.
- Bilkhu-Thompson, M. (2003). A process evaluation of a health care Balanced Scorecard. *Journal of Health Care Finance*, *30*(2), 37–64.
- Brennan, D. J., Robson, M. S., Murphy, M., & O'Herlihy, C. (2009). Comparative analysis of international cesarean delivery rates using 10-group classification identifies significant variation in spontaneous labor. *American Journal of Obstetrics and Gynecology*, *201*(308), 1–8.
- Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science*, *4*, 50. doi:10.1186/1748-5908-4-50.
- Dixon-Woods, M., Amalberti, R., Goodman, S., Bergman, B., & Glasziou, P. (2011). Problems and promises of innovation: Why healthcare needs to rethink its love/hate relationship with the new. *BMJ Quality and Safety*, *20*(1), i47–i51. doi:10.1136/bmjqs.2010.046227.
- Donabedian, A. (2005). Evaluating the Quality of Medical Care. *Milbank Quarterly*, *83*, 691–729.
- Fixsen, D.L., Naoom, S.F., Blasé, K.A., Friedman, R.M., Wallace, F. (2005). Implementation Research: A Synthesis of the Literature. In Louis de la Parte (Ed.). *Implementation Research: A Synthesis of the Literature*, University of South Florida, Tampa: Florida Mental Health Institute Publication. <http://ctndisseminationlibrary.org/PDF/nirmonograph.pdf>
- Ghaffarzadegan, N., Lyneis, J., & Richardson, G. P. (2011). How small system dynamics models can help the public policy process. *System Dynamics Review*, *27*(1), 22–44.
- Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of innovations in service organizations: Systematic review and recommendations. *Milbank Quarterly*, *82*, 581–629.
- Kaplan, R. S., & Norton, D. (1992). The Balanced Scorecard: Measures that drive performance. *Harvard Business Review*, *70*(1), 71–79.
- Klein, K. J., & Sorra, J. S. (1996). The challenge of innovation implementation, *Academy of Management Review*, *21*(4), 1055–1080.
- Istituto Superiore di Sanità Ministero della Salute (2010). Sistema nazionale per le Linee Guida. Linea guida 19. Taglio Cesareo: una scelta appropriata e consapevole. (National Institute of Health Care. Ministry of Health Care. National System for the Guidelines. Guideline 19. Cesarean Section: An appropriate and conscious choice). http://www.salute.gov.it/imgs/C_17_publications_1330_attachedfile.pdf.
- Li, H., Liu, J., & Blustein, J. (2013). Cesarean Delivery on Maternal Request. *JAMA*, *310*(9), 977–978.
- Li, H., Luo, S., Trasande, L., Hellerstein, S., Kang, C., Li, J., et al. (2017). Geographic variations and temporal trends in cesarean delivery rates in China, 2008–2014. *JAMA*, *317*(1), 69–76.
- Macfarlane, A. J., Blondel, B., Mohangoo, A. D., Cuttini, M., Nijhuis, J., Novak, Z., et al. (2016). Wide differences in mode of delivery within Europe: Risk-stratified analyses of aggregated routine data from the Euro-Peristat study. *BJOG*, *123*(4), 559–568.

- Nelson, E., Batalden, P., Godfrey, M., Headrick, L., Huber, T., Mohr, J., et al. (2002). Microsystems in health care: The essential building blocks of high performing systems. *The Joint Commission Journal on Quality Improvement*, 28(9), 472–493.
- Pollitt, C., & Bouckaert, G. (2004). *Public management reform: A comparative analysis* (2nd ed.). Oxford: Oxford University Press.
- Robson, M., Hartigan, L., & Murphy, M. (2013). Methods of achieving and maintaining an appropriate caesarean section rate. *Best Practice and Research Clinical Obstetrics and Gynaecology*, 27(2), 297–308.
- Triunfo, S., Ferrazzani, S., Lanzone, A., & Scambia, G. (2015). Identification of obstetric targets for reducing cesarean section rate using the Robson Ten Group Classification in a tertiary level hospital. *European Journal of Obstetrics, Gynecology, and Reproductive Biology*, 189, 91–95.
- Wildman, K., Blondel, B., Nijhuis, J., Defoort, P., & Bakoula, C. (2003). European indicators of health care during pregnancy, delivery and the postpartum period. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, 111, S53–S65.
- World Health Organization. (1985). Appropriate technology for birth. *The Lancet*, 2(8452), 436–437.
- World Health Organization. (2001). Antenatal care randomised trial for the evaluation of a new model of routine antenatal care. *The Lancet*, 357(9268), 1551–1564.
- Zelman, W., Pink, G., & Matthias, C. (2003). Use of the Balanced Scorecard in health care. *Journal of Health Care Finance*, 29(4), 1–16.
- Zizza, A., Tinelli, A., Malvasi, A., Barbone, E., Stark, M., De Donno, A., et al. (2011). Caesarean section in the world: a new ecological approach. *Journal of Preventive Medicine and Hygiene*, 52, 161–173.

Part II
Towards Outcome-Based Performance
Management: Experiences and Trends
from Different Countries

Chapter 5

The Scottish Government's System of Outcome-Based Performance Management: A Case Study of the National Performance Framework and Scotland Performs

Bobby Mackie

Abstract This chapter considers the Scottish Government's National Performance Framework as the overarching aspiration of the Scottish Government which in turns informs and guides public managers across Scotland's public services. The focus in the chapter is therefore on understanding this system of organisational performance management and the demands of outcome-based performance management on Scotland's public services and their management. This case study explores the research question on the efficacy of outcome-based organisational performance management in a government context and in particular the ways in which the Scottish Government has implemented its National Performance Framework and its system of communicating performance 'Scotland Performs'. This case study demonstrates that there are great difficulties in accommodating outcomes in systems of organisational performance management in government but nevertheless governments across the globe are pursuing outcome agendas. Organisational performance management systems in a government context must respond to this change by developing and effectively implementing comprehensive, output and outcome-focused, systems of organisational performance management.

Keywords Organisational performance management · Outcome-based · Performance measurement · Scottish Government · Scotland Performs

B. Mackie (✉)

School of Business and Enterprise, The University of the West of Scotland,
Hamilton Campus, Hamilton, UK
e-mail: bobby.mackie@uws.ac.uk

5.1 Introduction

This chapter considers the implementation and evaluation of ‘Scotland Performs’ which communicates via its website the Scottish Government’s achievements in relation to the aspirations set out in the National Performance Framework. The purpose of this chapter is to respond to Pollitt’s (2006) observation that there have been limited analyses of what elected members do with performance information and to Arnaboldi, Lapsley and Steccolini’s (2015) encouragement of researchers to undertake more nuanced research in this most difficult, complex, testing area for researchers and practitioners alike. The focus of this chapter is therefore on utilising a case study approach to explain and analyse the Scottish Government’s system of outcome-based performance management.

This case study explores the research question on the efficacy of organisational performance management in a government context and incorporates a review of contemporary literature on this topic. A case study approach to research involves an empirical investigation of a phenomenon within its real life context using multiple sources of evidence. Case studies can incorporate comprehensive descriptions of current managerial practices and this chapter describes the Scottish Government’s National Performance Framework and ‘Scotland Performs’. Case studies often conclude by making prescriptions for future action to enhance performance and also proscriptions on matters where mistakes/errors have been made. The approach adopted in this research is that of a single case incorporating responses to the following questions:

1. Have public organisations adopted performance measurement systems?
2. How are the measures used for decision making?
3. What are the main drivers relating to the adoption, use and effectiveness of performance measures in public administration?
4. How do performance management systems affect the relationship between policy-makers, public managers and external stakeholders?
5. In what circumstances to performance management systems predominantly have symbolic purposes?
6. What is the future of performance management in public organisations?

The case of ‘Scotland Performs’ can be classified as an intensive case (Saunders et al. 2012) as it contains a large amount of data on the practice of outcome-based performance management in a Government context. Data was obtained through a ‘Freedom of Information (FoI)’ request to the Scottish Government and through semi-structured interviews with senior Civil Servants in the Scottish Government and a range of public service managers. The interviews took place in 2016.

5.2 The Development of Outcome-Based Public Management

Baur (1966) noted a growing interest in social goals and indicators and there reporting on such subjects as the reduction in poverty, freedom from discrimination, social and political participation, civil liberties and the administration of justice, art and culture, employment and leisure, learning and education, health and well-being, the production of knowledge, the natural environment, the urban environment, and the mass media. Gross in the same text (Baur 1966, Chap. 3) also noted that the maturation of social accounting concepts will take many decades. Therefore interest in outcome-based performance management is not new and may be considered to be 'old wine in new bottles' but there has doubtless been an exponential growth in interest in outcome-based performance manage as a consequence of the range of global developments broadly classified as 'New Public Management (NPM)' and to the developments in communication and information technologies which facilitate the collection, analysis and dissemination of complex social and performance data. Governments and public organisations internationally have been changing their approach to management of public services. For many years there has been a focus on inputs, processes and outputs, and performance was largely assessed on how allocated budgets were spent and how processes were followed (Carter et al. 1993). There has been a shift in approach to enable governments to promote and measure progress in relation to 'well-being' and to consider this in terms of outcomes - or what makes a meaningful difference to the quality of people's lives.

5.3 The Scottish Government's National Performance Framework

The Scottish Government's National Performance Framework was created in 2007 and has changed the relationships between public sector organisations and the Scottish Government and between public sector organisations and funders acting on behalf of the Scottish Government. The emphasis in performance governance is on effective public policy implementation, performance measurement, accountability and value for money. In 2007 the Scottish Government commissioned a literature review on organisational performance management in a government context which provided an evidence basis for the Scottish Government's system of organisational performance management 'Scotland Performs' (Mackie 2008).

Pollitt (2001), Halligan (2007) and others suggest that global convergence is a consequence of political and economic aspirations for achieving particular outcomes and this is leading to a greater commitment on the part of governments to achieve sustainable results. The National Performance Framework of the Scottish Government is therefore not unique in its aspirations nor in its managerial implications.

According to the Scottish Government (2011), outcome based processes often promote localism, in the form of greater devolution of power and decision-making to local government and local partnerships. This enables services to better reflect local priorities and distinctive needs and circumstances. They focus on improving the effectiveness of partnership working, where agencies co-ordinate their policies and services towards the joint pursuit of shared outcomes. This in turn can foster greater trust and better relationships among public bodies, and improved scope for innovation. This approach also demands the adoption of underpinning cultures and systems to support them. Strong leadership is needed to provide authority and ensure momentum behind an outcome focus. Support is also required elsewhere in the system, including at middle-management levels, to build awareness and skills which enable outcomes-based principles to pervade throughout organisations.

Systems changes implied in an outcomes focuses approach include the development of new performance management and reporting arrangements, using performance information which allows progress towards outcomes to be measured. In this way, as well as providing a basis for performance monitoring and review, an outcome-based approach provides a potentially very powerful means of demonstrating how governments are addressing the needs and concerns of their citizens. The implementation of outcomes-based approaches necessitates multiple delivery partners and the role of the Scottish Government is to concentrate on providing leadership and direction, and to focus on strategic national priorities.

Following the election in May 2007 the Scottish National Party formed a minority Government in the Scottish Parliament and changed the collective term for the Government and its departments to “the Scottish Government” . In November 2007 the Scottish Government published a spending review containing a new national performance framework. The spending review contains five “strategic objectives” supporting delivery of the purpose and, in turn, these are supported by “national outcomes” which describe in more detail what the government wants to achieve over a 10 year period. Progress on these outcomes would be measured through “National Indicators and Targets”. The Scottish Government acknowledged the need for government to take a more strategic approach to target setting and set targets where the Scottish Government judge that it will be an incentive to delivery. Elsewhere in the spending review the Scottish Government established the direction of travel in which it expects indicators to move in the spending review period.

The focus of the Scottish Government's Purpose is on creating a more successful Scotland with opportunities for all to flourish. The Scottish Government believes that sustainable economic growth is the avenue through which the Scotland can achieve this and deliver a fairer, smarter, healthier, safer and greener society. Within the National Performance Framework, national wellbeing is covered through a wide range of social and environmental indicators and targets including mental wellbeing, income distribution and carbon emissions as well as economic growth (Scottish Government, 2008).

'Scotland Performs' is the Scottish Government's online tool for reporting on progress on overall delivery of its Purpose and National Outcomes. The Scottish Government is committed to the reform public services with a decisive shift towards prevention, greater collaboration, partnership working, transparency and workforce development. Excellent public services are essential for a productive and equitable society. The Scottish Government have formally recognised the strength of the public's commitment to Scotland's public services and believe that the quality of those services is the bedrock on which Scottish society and future prosperity depend (Scottish Government: <http://www.scotland.gov.uk>).

The Scottish Government has five objectives that underpin its core purpose—to create a more successful country, with opportunities for all of Scotland to flourish, through increasing sustainable economic growth:

Wealthier and fairer Scotland
Healthier
Safer and stronger
Smarter
Greener

Progress towards the Purpose is tracked by eight purpose targets and it is supported by 16 National Outcomes—describing the kind of Scotland the Scottish Government wants Scotland to be—and 55 National Indicators, covering key areas of health, justice, environment, economy, and education measure progress.



In December 2011, a National Outcome relating to older people was added to the National Performance Framework. The 16 National Outcomes are:

NATIONAL PERFORMANCE FRAMEWORK

THE GOVERNMENT'S PURPOSE
To focus government and public services on creating a more successful country, with opportunities for all of Scotland to flourish, through increasing sustainable economic growth

HIGH LEVEL TARGETS RELATING TO THE PURPOSE
Growth Productivity Participation Population Solidarity Cohesion Sustainability

STRATEGIC OBJECTIVES

WEALTHIER & FAIRER	SMARTER	HEALTHIER	SAFER & STRONGER	GREENER
NATIONAL OUTCOMES	We live in a Scotland that is the most attractive place for doing business in Europe			
	We realise our full economic potential with more and better employment opportunities for our people			
	We are better educated, more skilled and more successful, renowned for our research and innovation			
	Our young people are successful learners, confident individuals, effective contributors and responsible citizens			
	Our children have the best start in life and are ready to succeed			
	We live longer, healthier lives			
	We have tackled the significant inequalities in Scottish society			
	We have improved the life chances for children, young people and families at risk			
	We live our lives safe from crime, disorder and danger			
	We live in well-designed, sustainable places where we are able to access the amenities and services we need			
	We have strong, resilient and supportive communities where people take responsibility for their own actions and how they affect others			
	We value and enjoy our built and natural environment and protect it and enhance it for future generations			
	We take pride in a strong, fair and inclusive national identity			
	We reduce the local and global environmental impact of our consumption and production			
Our people are able to maintain their independence as they get older and are able to access appropriate support when they need it				
Our public services are high quality, continually improving, efficient and responsive to local people's needs				
HOW ARE WE DOING? Visit www.scotlandperforms.com to track latest progress				

March 2016

5.4 National Indicators

The National Outcomes are directly linked to a set of National Indicators (now 55). Scotland Performs offers accountability based on national priorities set out in the National Performance Framework. Individuals can judge for themselves how Scotland is progressing by accessing Scotland Performs via the Scottish Government website. Scotland Performs measures how Scotland is progressing through 'direction of travel' arrows on the 'Performance at a Glance' page which indicate whether performance is improving, worsening or maintaining. Assessments of progress are regularly updated from the latest evidence and each has explanatory notes attached.



There are ten guiding principles for Scotland Performs:


- Openness and transparency.
- Accountability and responsibility.
- Objectivity.
- Independent assessment.
- Dynamic site: real data, real time.
- Accessibility 24/7.
- Simplicity and clarity.
- Credibility to Parliament and the wider public.
- Shared responsibility for outcomes-based performance (with our partners).
- Sharpening focus—driving improvement.


National Performance Framework – Measurement Set			
Increase Scotland's Economic Growth	Improve Productivity	Improve Economic Participation	Increase Population Growth
PURPOSE TARGETS			
Population – Increase Healthy Life Expectancy	Solidarity – Reduce Income Inequality	Cohesion – Reduce Inequalities in Economic Participation across Scotland	Sustainability – Reduce Greenhouse Gas Emissions
<p>NATIONAL INDICATORS</p> <ul style="list-style-type: none"> Increase the number of businesses Increase exports Improve digital infrastructure Reduce traffic congestion Improve Scotland's reputation Increase research and development spending Improve knowledge exchange from university research Improve the skill profile of the population Reduce underemployment Reduce the proportion of employees earning less than the Living Wage Reduce the pay gap Increase the proportion of pre-school centres receiving positive inspection report Increase the proportion of schools receiving positive inspection reports Improve levels of educational attainment Increase the proportion of young people in learning, training or work Increase the proportion of graduates in positive destinations Improve children's services Improve children's dental health Increase the proportion of babies with a healthy birth weight Increase the proportion of healthy weight children Increase physical activity Improve self-assessed general health Improve mental wellbeing Reduce premature mortality Improve end of life care Improve support for people with care needs Reduce emergency admissions to hospital 	<ul style="list-style-type: none"> Improve the quality of healthcare experience Reduce the percentage of adults who smoke Reduce alcohol related hospital admissions Reduce the number of individuals with problem drug use Improve people's perceptions about the crime rate in their area Reduce reconviction rates Reduce crime victimisation rates Reduce deaths on Scotland's roads Improve people's perceptions of the quality of public services Improve the responsiveness of public services Reduce the proportion of individuals living in poverty Reduce children's deprivation Improve access to suitable housing options for those in housing need Increase the number of new homes Widen use of the Internet Improve people's perceptions of their neighbourhood Increase cultural engagement Improve the state of Scotland's historic sites Improve access to local greenspace Increase people's use of Scotland's outdoors Improve the condition of protected nature sites Increase the abundance of terrestrial breeding birds: biodiversity Increase natural capital Improve the state of Scotland's marine environment Reduce Scotland's carbon footprint Increase the proportion of journeys to work made by public or active transport Reduce waste generated Increase renewable electricity production 	<p>NATIONAL INDICATORS</p>	
<p>Visit www.scotlandperforms.com to track latest progress</p>			
<p>March 2016</p>			






















































Key to arrows

Key to arrows

 Performance Improving
 Worsening

 Performance Maintaining

 Performance Worsening

- | | |
|---|--|
|  Increase the number of businesses |  Improve the quality of healthcare experience |
|  Increase exports |  Reduce the percentage of adults who smoke |
|  Improve digital infrastructure |  Reduce alcohol related hospital admissions |
|  Reduce traffic congestion |  Reduce the number of individuals with problem drug use |
|  Improve Scotland's reputation |  Improve people's perceptions about the crime rate in their area |
|  Increase research and development spending |  Reduce reconviction rates |
|  Improve knowledge exchange from university research |  Reduce crime victimisation rates |
|  Improve the skill profile of the population |  Reduce deaths on Scotland's roads |
|  Reduce underemployment |  Improve people's perceptions of the quality of public services |
|  Reduce the proportion of employees earning less than the Living Wage |  Improve the responsiveness of public services |
|  Reduce the pay gap |  Reduce the proportion of individuals living in poverty |
|  Increase the proportion of pre-school centres receiving positive inspection reports |  Reduce children's deprivation |
|  Increase the proportion of schools receiving positive inspection reports |  Improve access to suitable housing options for those in housing need |
|  Improve levels of educational attainment |  Increase the number of new homes |
|  Increase the proportion of young people in learning, training or work |  Widen use of the Internet |
|  Increase the proportion of graduates in positive destinations |  Improve people's perceptions of their neighbourhood |
|  Improve children's services |  Increase cultural engagement |
|  Improve children's dental health |  Improve the state of Scotland's historic sites |
|  Increase the proportion of babies with a healthy birth weight |  Improve access to local green space |
|  Increase the proportion of healthy weight children |  Increase people's use of Scotland's outdoors |
|  Increase physical activity |  Improve the condition of protected nature sites |
|  Improve self-assessed general health |  Increase the abundance of terrestrial breeding birds: biodiversity |
|  Improve mental wellbeing |  Increase natural capital |
|  Reduce premature mortality |  Improve the state of Scotland's marine environment |
|  Improve end of life care |  Reduce Scotland's carbon footprint |
|  Improve support for people with care needs |  Increase the proportion of journeys to work made by public or active transport |
|  Reduce emergency admissions to hospital |  Reduce waste generated |
| |  Increase renewable electricity production |

The National Performance Framework provides a clear vision for the kind of Scotland the Scottish Government wants to see. The premise is that

outcome-focused working helps public services and other key contributors to work together effectively to tackle Scotland's key long-term economic, social and environmental challenges. The Scottish Government believe that making the best use of Scotland's collective resources will tackle the country's most difficult problems such as alcohol misuse and health inequalities and really make a difference to the quality of life and experience for the people of Scotland.

An updated National Performance Framework (NPF) indicator set was published on 11 March 2016. There are no structural changes to the NPF itself. The Government's Purpose along with the Strategic Objectives and National Outcomes remained unchanged. Changes have been made to the Productivity, Solidarity and Sustainability Purpose Targets following consideration of changing environmental circumstances requiring target modifications (Scottish Government, 2016c).

A key feature of the National Outcomes is their dependence on partnership working. The Scottish Government concluded a revised Concordat with Scottish local authorities in late 2007 which emphasises the significant part local government has to play in promoting the achievement of the National Outcomes. Central to this revised Concordat is the introduction of 32 Single Outcome Agreements (SOAs) between Scottish local authorities and the Scottish Government.

The Concordat agreed between the Scottish Government and the Convention of Scottish Local Authorities (COSLA) sets out the terms of a new relationship between national and local government, based on mutual respect and partnership. This new relationship is represented by a package of measures endorsed by the Scottish Government and the Convention of Scottish Local Authorities (COSLA) and which both parties believe will, over time, lead to significant benefits for users of local services across Scotland. A key element of the Concordat has been the move to create Single Outcome Agreements (SOAs) between all 32 local authorities in Scotland and the Scottish Government. The SOAs are to be based on the set of national outcomes and, under a common framework, local outcomes, to take account of local priorities (Scottish Government, 2009).

A high level steering group (HLSG) established by the Concordat is overseeing the development and implementation of the SOAs. The HLSG comprises senior representation from the Scottish Government, COSLA, the Society of Local Authority Chief Executives (SOLACE), Audit Scotland, and the Improvement Service; the HLSG is chaired by COSLA. All direct engagement between the Scottish Government and the local authorities on developing their SOAs is being managed through the Scottish Government Implementation Group (SGIG). The SGIG comprises 11 Scottish Government Directors, each of whom has been assigned either one or two National Outcomes to lead on, a policy home area and, in some cases, further affiliated areas. This aligning of expertise allows the Group collectively to form a view across all aspects of Government policy.

For the purposes of co-ordinating the liaison with each local authority, 9 of the 11 Directors on the Group has additionally been assigned up to 4 councils to work with and takes the lead in any discussion on the content and development of SOAs with those particular councils. Each Director has a Support Team to assist in the management of their relationship with each council.

Each SOA expresses the joint commitments between the local authority, its community planning partners and the Scottish Government to the delivery of an agreed set of outcomes. Most councils have identified actions which they have requested the Scottish Government to take to support the delivery of the outcomes. Each party to the agreement is mutually accountable for the delivery of the agreed outcomes and will jointly take ownership and responsibility for their respective contributions. They will also be able to hold each other to account for the delivery of specific commitments they make to enable the delivery of outcomes.

The Concordat states that the Scottish Government will step back from micro-managing how councils deliver services for their communities, while supporting their delivery of their SOAs. A corollary of that is an increased onus on councils to ensure that they are able to design, operate and deliver services in a way that supports better outcomes effectively. Councils are therefore responsible for sound governance and for applying robust performance management practices and the Scottish Government will ensure that its NDPBs and agencies align their practices to these arrangements, for the joint delivery of agreed outcomes (Scottish Government, 2009).

The Local Government in Scotland Act 2003 places a statutory duty of Best Value on all councils. Councils have their own performance management arrangements and the Scottish Government will not prescribe use of a particular performance management system. However, in general terms councils will need to make sure that performance management systems collect relevant information to report on their delivery of agreed outcomes. Councils are expected to use the best available indicators to track and support delivery of their outcomes and these may be specific to their area, rather than using less relevant indicators simply to provide national comparability. Councils should also have mechanisms in place to assess and act appropriately upon this information and other evidence of performance against outcomes. Single Outcome Agreements (SOAs) have now been extended to all public bodies in Scotland and to all colleges and universities.

The Community Empowerment (Scotland) Act 2015 continues a commitment to the outcomes approach to government. This means that the focus on achieving goals that improve the wellbeing and quality of life of the people of Scotland will continue to be a priority for the present administration. A vision for Scotland will be developed by the Scottish Government in consultation with the people of Scotland and progress towards this measured so the Scottish Government will know whether the aspirations contained in the vision are being realised.

Specifically, the Act places a duty on the Scottish Ministers to consult on, develop and publish a set of national outcomes targets for Scotland. The Scottish Ministers must also regularly and publicly report on progress towards these outcomes and review them at least every five years. When setting the national outcomes targets, the Scottish Ministers must have regard to the reduction of inequalities of outcomes which result from socio-economic disadvantage.

The National Performance Framework (NPF) provides a strategic direction for policy making in the public sector, and provides a clear direction to move to outcomes-based policy making. This outcomes-based approach is reflected across Government policy and in strategic policy documents. This can be evidenced by

rhetoric contained in the Scottish Government Programme for Government (September 2016) and the Scottish Budget: Draft Budget 2017–2018 (December 2016). Scotland is one of the first countries to publically sign up to the United Nations Sustainable Development Goals. The NPF will be one of the measures used to monitor Scotland’s progress towards these goals.

5.5 The Global Development of Organisational Performance Management

Organisational Performance Management (OPM) in a public service is the managerial activity necessary to promote well-performing policy management and service delivery (United States Government Accountability Office (USGAO), 2017). A desire for improved performance in public sector organisations has resulted in a results-orientation and a cost consciousness in a range of Organisation for Economic Co-operation (“OECD”) countries (OECD 1997, 2015)). Performance management systems often utilise a performance information system that can be audited and is related to financial management and policy cycles (Treasury Board of Canada Secretariat, 2003) although this approach can lead to an unacceptable administrative burden on governmental bodies. Organisational performance management in a government context concerns monitoring the success of public policy, programmes or projects in achieving their objectives and in securing the expected benefits (World Bank, 2015).

Organisational performance management in a public service context is therefore the activities of government or its agencies in planning, implementing, reviewing, evaluating and reporting, the effectiveness of its policies, programmes and projects. The key purpose of organisational performance management is to introduce systematic controls in the management process to guide and regulate the activities of an organisation or any of its parts, by means of management judgement, decision, and action for the purposes of attaining agreed objectives.

5.6 The Implementation of the National Performance Framework

National Performance Frameworks (NPFs) enable government to drive, monitor and assess progress towards achieving their overarching national objectives. NPFs also provide an accountability framework through which parliaments and civil society can measure the effectiveness of government action (OECD, 2015).

Half of OECD member countries have a NPF in place, although the types of framework differ substantially. In some countries the NPF is developed and monitored by the Ministry of Finance, while in other countries it is developed by the statistical agency. There are also some countries where the NPF is a joint project by government

departments and the statistical agency. Who drives the framework can affect the extent to which there is political buy-in as well as its purpose. It can also determine whether or not there are specific targets attached to indicators (OECD, 2015).

In Scotland, public services, working with partners, play a key role in delivering the Scottish Government's Purpose and National Outcomes. At a local level, Community Planning Partnerships support the delivery of the National Performance Framework (NPF) through individual agreements between public services and their partners in delivery.

The Deputy First Minister chairs a Round Table Group with representation from all political parties in the Scottish Parliament and representatives of Scotland's public services, third sector organisations and academics to monitor the Scotland Performs data and its development. The Scottish Parliament Committees receive regular updates from Scotland Performs to enhance the evidence basis for public policy and the performance management of the National Performance Framework. The Scotland Performs Technical Advisory Group (SPTAG) advises on the suite of national indicators and on the data collection and data presentation.

Organisational performance management in a government context can serve two distinct functions:

- **Intra-organisational performance management:** To ensure that there are appropriate internal controls to monitor the extent to which the organisation (and its sub-units) is achieving what it is supposed to achieve. This requires the organisational management to periodically review and evaluate performance standards attained and performance trajectories, taking corrective action as appropriate where deviations from the desired standards are detected (Mackie, 2013).
- **Extra-organisational performance management:** To communicate performance for the purposes of governance and accountability to organisational stakeholders including Government, funding bodies, audit agencies and the wider public (Ibid).

There is no legislative requirement for an organisation to have an intra-organisational performance management system. Organizations need to know where they are, where they are going and how to manage the changes. Managers in these organizations need to know where their roles fit in relation to the whole and how they can contribute to strategic developments and changes

There is a widely accepted belief that having clarity of purpose and the means to monitor progress towards goal attainment does promote a performance focus in organisations (public and private) and as such is more likely to achieve enhanced organisational performance levels. There is no guarantee of enhanced performance levels as performance achieved depends on a range of variables only one of which is clarity of direction. There are requirements, often statutory, for public sector organisations to maintain high standards of corporate governance, accountability and public reporting. This requires systems of extra-organisational performance management.

Extra-organisational performance management involves controlling organisational resources and activities to ensure that they are contributing to organisational effectiveness and to ensure that the organisation is not experiencing strategic drift. Strategic drift occurs when the reality of organisational performance is inconsistent with planned levels of performance. The management of organisational performance is an activity of senior management as they are most likely to be held accountable by politicians and other stakeholders for performance levels achieved and there is growing evidence of organisational performance being a core feature of systems of people performance management such as annual performance planning and therefore sensitivity to organisational performance has become a key activity of managers at all levels.

If there is a major deviation between a national performance framework's planned and actual performance detected by performance measures then governments must consider adjusting performance or modifying plans accepting that in many circumstances actual performance levels may be outwith the control of governmental and managerial activity and performance gaps persist over time. It is therefore important for governments to make sure that all levels of public service delivery are in touch with each other and work together to do their best to deliver the governments aims.

Organisational performance management in a public service context should with political priorities. The OECD (2015) has identified five recurring themes of systems of organisational performance management in a government context: economy; environment; education; health; and society. The Scottish Government's National Performance Framework (NPF) derived its initial priorities from the Scottish National Party Manifesto for the 2007 Scottish Parliamentary Elections which put the Scottish National Party in power for the first time since devolution in 1999. The revision to the NPF in 2011 and 2016 were driven by a process combining political priorities and public consultation (Scottish Government, 2011 and 2016a, b,c).

Public sector organisations that prioritise well incorporate the following factors:

- evidence from stakeholders and the public which has been used to establish aims and priorities;
- politicians who are involved in setting strategic aims and in ranking them;
- aims and priorities, and their relative importance, that are clear and underpin the vision and strategy;
- resources that are linked to aims and priorities;
- aims and priorities which have been communicated internally and externally;
- aims and priorities are cascaded down to individual actions;
- the existence of systems to support monitoring of this activity'
- the use of clear milestones and measures to underpin the political vision'
- partners' priorities and plans to reflect political priorities and vice versa'
- priorities that are reviewed at appropriate intervals to reflect changing demands and current progress.

(IDeA 2004)

Organisational performance management in a public service is the managerial activity necessary to promote well-performing policy management and service delivery. A desire for improved performance in public sector organisations has resulted in a results-orientation and a cost consciousness in a range of OECD countries (OECD 1997, p. 8)

Research shows that there are multiple tools and techniques being used globally in public sector performance management (World Bank, 2007 and 2015). There is clearly a tendency towards colour charts and diagrams providing a snapshot of the current status of actual performance against planned performance using information which is as close as possible to real time. These tools have most significance where the focus of the performance relates to service inputs, process and outputs and is disaggregated by service and residential area. Corrective action can be introduced relatively quickly and the expectation is that the corrective action will have a short term impact. Difficulties arise when the focus of the performance is on national level outcomes pursued over the longer term. However 'dashboard' type communication of performance status does have a place in government performance management systems (OECD, 2015).

Government's success or otherwise in achieving policy outcomes will be as a consequence of aggregated performance over a period of time. If sub-units and programmes are achieving their objectives then they will contribute to the attainment of organisational objectives. This is an approach developed from Drucker's (1955) 'Management by Objectives' (MBO) concept. There are difficulties in the process by which long term outcomes are translated to shorter term targets and subsequent cascading of targets horizontally and laterally (through organisational hierarchies and between organisations). But the process of attempting to translate outcomes into process and output targets can be of value in itself as it can lead to a better understanding on the part of public service managers of the fundamental purposes of their roles and the ultimate consequences of their performance (Mackie, 2013, p. 64).

Many academics remain critical of managerialism in the public sector but others believe it is better to approach the task of public service provision with greater clarity of desired future and a well-developed sense of direction. Tools of performance management must contribute to more effective public management but there is much research and evaluation required to determine the ways in which generic management approaches can be adapted for the distinctiveness of public services organisational performance management.

National cultures exert influence over organisational and governmental practices in many countries (Hofstede, 2001) and there is clear evidence of global convergence in relation to organisational performance management in a government context. According to Pollitt (2001, p 943):

Many benefits flow to many players from a situation in which there is a dominant, but loosely-specified set of reform ideas which apparently can be applied to a very wide range of public sector contexts.

There is therefore an expectation that changes in organisational performance management in one system of government would be replicated in other governments at least in the short term. The OECD (2007, p. 19) has noted that 75% of OECD countries have introduced a new initiative on performance management in government. Despite apparent convergence, there remains diversity across countries and differences within countries reflecting governmental policy priorities. The triggers for change are commonly financial crises, pressure to reduce public expenditures and changes in political administration. The objectives of the reforms focus on:

- Budgetary priorities of expenditure control and improving allocative efficiency and productive efficiency;
- Results based management and improving public sector service delivery, efficiency and performance; and
- Improving accountability to politicians and the public.

(OECD, 2007:24–25)

The Scottish Government consults with its partners in the delivery of public services to develop a common understanding of the Government's aims and to identify which activities make a real difference to the well-being of Scotland's population. In addition there is an ongoing dialogue on the meaning of outcomes and the contributions necessary from the Government partners in delivery. The Scottish Government need to promote alignment between the activities of those who deliver public services and the Government's aspirations as expressed in the National Performance Framework. In order to promote this alignment Senior Civil Servants (Directors) are allocated, as part of their duties and responsibilities, the monitoring of progress against the 16 National Outcomes through direct engagement with public service delivery partners. The Scottish government can exercise more control in certain areas of public service delivery but need to take heed of subsidiarity and local priorities. The overall objective of the National Performance Framework is to achieve a more focused, evidence based approach to the planning and management of governmental activity through cross public service dialogue leading to the development of an outcomes culture which permeates public service management in Scotland.

5.7 Key Questions on the Efficacy of Outcome-Based Performance Management Systems

Based on the experience of Scotland Performs and the Scottish Government's National Performance Framework six key questions can be addressed:

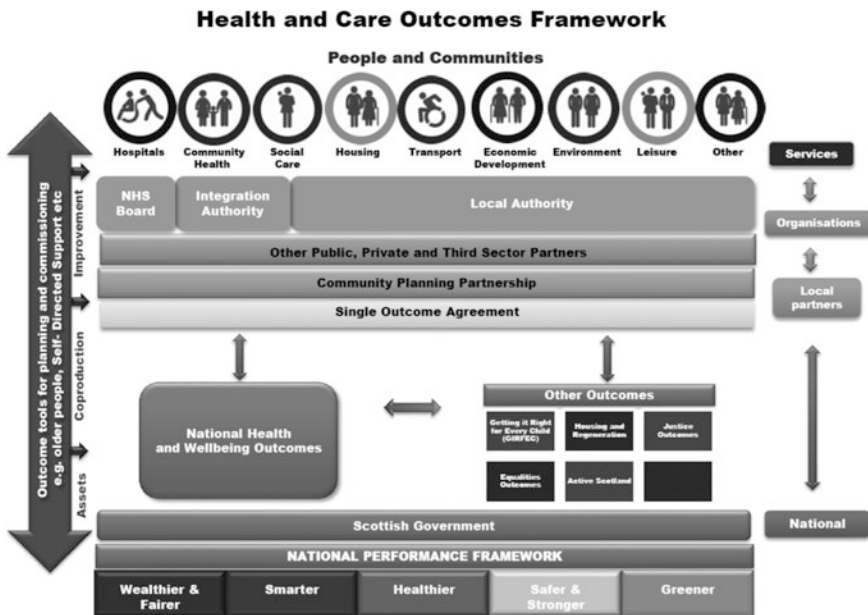
1. Have public organisations adopted performance measurement systems?

The research interviews conducted across Scotland's public services reveal that as a direct consequences on the introduction of the Scottish Government's National Performance Framework the vast majority of public organisations in Scotland have adopted performance measurement systems for internal (performance management) and external use (accountability and stakeholder communication). Such systems did not exist in Scottish public service organisations prior to 2008 (Mackie, 2013).

2. How are the measures used for decision making?

Performance indicators (PIs) are the measures and can be defined as data for intra and extra organisational use mainly in a quantified form on aspects of organisational input, activity, output and outcome; that focuses on the actual past, the present and the projected future of an organisation as an aid to assessing the extent to which the organisation is pursuing and attaining its mission and objectives in an effective and efficient manner (Mackie, 2005).

Scotland Performs submits the latest data available to the Scottish Government and to Committees of the Scottish Parliament. The measures, as a consequence, become an evidence basis and an input the political decision-making process. At sub-national levels there are other performance frameworks which link to the National Performance Framework and provide guidance to service providers in specific public service areas such as Health and Care as illustrated in the diagram below:



The Scottish Government have recognised weaknesses in this initial model and in the Draft Budget 2017-18 a further series of performance related reforms are planned:

Individually and collectively these reforms, like those that have been implemented in the past year, seek to improve outcomes for people at every life stage. By focusing on outcomes we hope to develop and deploy the public service workforce in a way that establishes a truly preventative culture, one which forges deeper relationships with local people and is more open and responsive to what communities most value.

(Scottish Government, 2016b)

3. What are the main drivers in relation to the adoption, use and effectiveness of performance measures in public administration?

Performance measures in public administration can be used in three main ways. The first way is as a tool of performance governance. This is a ‘top-down’ driver where government set out their national performance frameworks and the expectation is that public services will interpret the national framework and utilise it to guide organisational activity. Governments monitor the performance of public services in relation to the extent to which public services align their strategic aspirations to the national framework and in relation to the results public services are achieving.

The second way is as a tool of performance management whereby public service managers at various levels use performance measures as targets (pre-controls), as tools for monitoring progress and performance trajectories (concurrent controls) and as tools of evaluation, review and reporting (post-controls). The driver here is managerial effectiveness.

The third way is as marketplace surrogates where measures are used to communicate to stakeholders and the public on aspects of public service performance standards attained. The drivers here are accountability and transparency.

4. How do performance management systems affect the relationship between policy-makers, public managers and external stakeholders?

Utilising the three ways in which performance related information can be used: performance governance, performance management and stakeholder communication, relationship can be affected in different ways. Performance management systems can enhance performance governance making policy-makers (politicians) better informed and more powerful. In this situation, public managers are more accountable to policy-makers for their performance at an organisational level. In relation to performance management, public managers should be better equipped to communicate performance standards achieved to both policy-makers and external stakeholders. In addition they should be better informed about their organisation’s performance and therefore in a better position to make good decisions. Stakeholder communication empowers stakeholders and promotes the accountability of both public managers and policy-makers to the public and to the electorate.

5. In what circumstances to performance management systems predominantly have symbolic purposes?

This case study illustrates that many of the Scottish Government's National Indicators evidence low levels of public interest as measured through 'hits' on the Scotland Performs website. This may be an indication that the Scottish Government has a desire to show the diversity of their concerns for the 'well-being' of Scotland but in reality they evidence limited activity in relation to some of the National Indicators. Some National Indicators are clearly more important to external stakeholders than others. The number of hits on the Scotland Performs website gives an indication of public interest and the Scottish Government must be sensitive to public interest to ensure that it maintains popular support. The top indicators as measured by the number of hits (in rank order) on the Scotland Performs website (2012–15) are as follows:

Adults who smoke
Deaths on Scotland's roads
Physical activity
Skill profile
Alcohol related hospital admissions
Emergency admissions to hospital
Number of businesses
Use of the Internet
Problem drug use
Healthy birth weight
Mental well being

Politicians are taking note of public interest in particular issues many of the issues are incorporated in the Scottish Government's Programme for Government (2016a, b). It is possible to identify who is making the hits on the website and the number of hits does indicate general interest from a range of stakeholders in particular indicators.

6. What is the future of performance management in public organisations?

The Scottish Government's National Performance Framework is here to stay for the next 5 years at least given the return to power of the Scottish National Party and the incorporation into their legislative proposals of many of the most popular policy issues as evidenced by the number of 'hits' on the Scotland Performs website.

In addition the Scottish Government's National Performance Framework and Scotland Performs according to senior Civil Servants have achieved international recognition as representing good practice in organisational performance management in a government context (see OECD, 2015).

5.8 Conclusion

Organisational performance management literature identified the core elements of organisational performance management in a government context and has confirmed that it is global development of the government. Reports by the World Bank (2007), the US Government Accountability Office (USGAO, 2017) and the OECD (2007) identify the pervasive characteristics of global organisational performance management in a government context. According to these influential organisations, comprehensive systems of organisational performance management in a government context should be modelled on these pervasive characteristics:

- High level public policy aspirations expressed as outcomes
- Strategic business plan
- Performance measurement tools and techniques
- Targets
- Implementation
- Monitoring
- Measuring results
- Verification
- Communication
- Review and evaluation
- Continuous sensitivity
- Commitment

However, the premise underpinning these answers is that performance improves in part as a consequence of an holistic outcome-based organisational performance management system and there is transparent evidence of strategic fit between public policy objectives and priorities, and the progress towards these objectives and priorities made by governmental and public services' performance. The difficulties of effective policy implementation and the problematic consequences of performance measurement systems will always apply and an organisational performance management system is no guarantee of policy success. Policies fail because of bad policy, bad execution or bad luck and there may be some sense that governments have to introduce systems of organisational performance management not because of their potential benefits but because other governments are introducing such systems and the rhetoric of an organisational performance management system in government may be sufficient to appease the public. The reality may be that the policy of organisational performance management becomes a substitute for action. This approach may be useful when confronted by a problem, which is difficult to address (low tractability), and policy implementation activity may result in no tangible signs of improvement. The policy as a statement of intent is a substitute for action and the organisational performance management system is never effectively implemented.

Outcome-based organisational performance management in a government context has multiple objectives some of which focus on governmental effectiveness and

others on responsiveness. In some cases the system contains more symbolic rhetoric so to project an indication of action when in fact there is limited chance of performance improving as the issue is more complex with no easy answer (wicked problems). In recent years governments globally have set out clear public policy objectives in terms of outcomes and are committed to the pursuit and attainments of these societal outcomes. Governments can be held accountable periodically to their electorates for their performance including the extent to which they have made progress towards the attainment of these outcomes. Much depends on the extent to which performance enters the public domain and is addressed by politicians (particularly those in opposition), the media, key stakeholders and the general public. The key to public debate is access and the accuracy of the performance information made available. If it is accessible and understandable then the systems offers potential for its use as a tool of accountability. There have been recent examples of the Scottish Government having to account for criticism of the performance of the school education system and this has led to education being a priority in the Programme for Government (2016) and in the Draft Budget 2017–18. This is clear evidence of the potential of the organisational performance management system to impact on government policy (Scottish Government 2016a and b).

Outcome-based organisational performance management in a government context will only be sustainable where it achieves its key objectives of enhancing the performance of governments in the attainment of its policy objectives and keeping the electorate and key stakeholders informed of the evaluations of the outputs and outcomes of such approaches. Academics tend to adopt a critical perspective where they are proscriptive about (i.e. critique) government managerialism and its alleged preoccupation with measures and targets. The rationale for so doing is the absence of empirical data confirming that there have been improvements in both outputs and outcomes as a consequence of an initiative which incorporates elements of organisational performance management. There are few examples of prescription by academics (see Gao, 2015) perhaps as a consequence of their limited or non-existent experience in a public management capacity. Academics, whose backgrounds lie outwith the management disciplines, cannot be expected to be advocates of particular organisational performance management systems. As a consequence much of the academic writings are from individuals with a social science and/or politics background and, although they provide excellent objectivity in their critiques, fail to incorporate recommendations that may lead to continuous improvement in organisational performance management in a government context. The identification of dysfunctional consequences abound but there is rarely a prescription of good practice to inform practitioner and organisational learning (see Smith, 1995). This case study of the Scottish Government's National Performance Framework and Scotland Performs provides an illustration of rationality and some success and can therefore be predominantly prescriptive in its advocacy of the outcomes focused approaches adopted by the Scottish Government.

Organisational performance management systems are high on government agendas worldwide and it must be assumed that there is global consensus within governments over the potential merits of such systems. Outcome-based

organisational performance management in government has the potential to enhance the effectiveness of public policy implementation but will only succeed where it is effectively implemented and applied holistically both within government and across governmental activity. Outcome-based performance management in government is globally fashionable but it may become a transient hobby (a fad) unless it is effectively implemented and continuously responsive to internal and external challenges. Politicians and public managers can promote the sustainability of outcome-based performance management but the extent to which it becomes embedded will depend on the extent to which outcome-based performance management performs.

There has been a shift from input controls to output controls and in more recent times, to outcome controls but this has not yet resulted in greater flexibility and looser control within organisations. Public sector organisations globally appear to be overwhelmed by forms of performance monitoring including scrutinies, audits, performance review systems, peer assessments, appraisals, statistical returns, etc. As a consequence of this there appears to be a contradiction in the role of outcome-based performance management in public management. Outcome-based performance management can be a tool of hands-off governance or it can support a rational-systems model of top-down control.

This case demonstrates that there are great difficulties in accommodating outcomes in systems of organisational performance management in government but nevertheless governments across the globe are pursuing outcome agendas. Organisational performance management systems in a government context must respond to this change by developing and effectively implementing comprehensive, outcome-focused, systems of organisational performance management. The Scottish Government National Performance Framework and Scotland Performs represent an innovative approach to outcome-based organisational performance management in a government context and as such are worthy of further research to ascertain the extent to which sustainable improvements in government performance can in-part be attributed to such systems.

Potential problems may arise when policy outcomes are not achieved or there are clearly gaps between desired performance levels and performance levels attained. Ideally, the organisational performance management system would cover all of government (holistic), the information would be as close to “real time” as possible and all of the information relating to the performance achieved would be in the public domain with well-presented summaries appearing in the form of traffic lights at regular intervals on a dedicated website such as ‘Scotland Performs’. Policy failure is attributable to bad policy, bad execution or bad luck. Bad policy can be addressed through better policy formulation. Bad execution can be addressed through better policy implementation and better organisational performance management. Bad luck can be partially addressed by better organisational performance management including risk identification, assessment and management. However, what do politicians do when there is apparent policy failure and that policy failure is evidenced by performance data in the public domain? What happens when an apparent policy failure appears prior to an election?

If governments are committed to the principles of outcome-based organisational performance management in a government context they must be prepared to address both the good news and the bad news. In saying that, the early evidence of comprehensive organisational performance management systems in the Scottish Government lead to the conclusion that it appears to have the potential to enhance performance, responsiveness and empowers citizens as evidenced by the Scottish Government's Programme for Government 2016 and the Draft Budget 2017–18 and as such it must be a positive development in public management.

It is too early to conclude that the impact of outcome-based organisational performance management systems in government has been a success as whole-of-government systems have only been introduced globally over the past 15 years. At sub-governmental levels the evidence (outwith the United Kingdom) is that such performance management systems have contributed to enhanced public service performance albeit that these systems have largely focused to date on municipal/local government and evaluations have tended to focus on inputs, process and output enhancement and not on outcomes achieved.

There is evidence of utility if the system is organisation-wide and linked to strategic planning and budgetary management systems. The organisational performance management systems act as concurrent controls providing essential management information as an aid to decision-making at all levels. Benefits relate to improved organisational performance, better management, better stakeholder communication and better relations (both internally and externally). Outcome-based organisational performance management systems in a government context have potential for multiple beneficiaries: politicians, civil servants, others involved in service delivery; other stakeholders; and the ordinary citizens. Public services in Scotland are now required to ensure the alignments of their key plans and strategies with those of its other partners engaged in public service delivery.

The extent to which the National Performance Framework remains intact will to some extent depend on the performance of the Scottish Government. There is an opportunity to do much better in managing organisational performance in Scottish Government but there is also a danger in the emerging proliferation of public service organisational performance management systems as they may develop in ways which are incompatible with the Scottish Government's aspirations.

The policy implementation framework for the National Performance Framework has established the need for public service organisations to demonstrate linkages between their planned activities and the 'Purpose Targets' and 'National Indicators' commonly through the development of detailed commitments from public services describing the ways in which they will align their activities to the National Performance Framework and in particular to promote progress on the National Indicators. This development over the past ten years contains both 'top down' and 'bottom up' elements in that the Scottish Government wish to exert influence (if not control) over the results of activity financed through public expenditure funds yet at the same encourage local decision making through partnership processes.

The approach adopted in Scotland is not unique but is significantly differentiated from the systems of other countries through the National Performance Framework

and Scotland Performs. In order to better understand outcome-based organisational performance management in public bodies, academics need to consider existing systems and evaluate their strengths and weaknesses. This case study creates an opportunity to examine a system which has been in existence for almost ten years thus providing a concrete experience of organisational performance management in public bodies.

This case study demonstrates that it is possible to introduce comprehensive systems of outcomes-focused organisational performance management in government and governments across the globe are pursuing outcome agendas (OECD, 2015). Organisational performance management systems in a government context must continue to respond to this change by developing and effectively implementing comprehensive, outcome-focused, systems of organisational performance management. The Scottish Government have made progress in developing and encouraging of an outcomes-focused culture in Scotland's public services by promoting the alignment of the National Performance Framework to other service focused frameworks and through Local Outcomes Improvement Plans generated by Community Planning Partnerships across Scotland. This approach promotes direct linkages between Scotland's public services and the National Performance Framework in an effort to ensure that every public body is aware of the needs to make an appropriate contribution to the National Outcomes as well as delivering their own statutory and permissive functions and services to the people of Scotland.

References and Selective Bibliography

- Arnaboldi, M., Lapsley, I., & Steccolini, I. (2015). Performance management in the public sector: The ultimate challenge. *Financial Accountability and Management*, 31(1), 1–22.
- Baur, R. A. (Ed.). (1966). *Social indicators*. Cambridge: The M.I.T. Press.
- Carter, N., Klein, R., & Day, P. (1993). *How organisations measure success: The use of performance indicators in government*. London: Routledge.
- Drucker, P. F. (1955). *The practice of management*. London: Routledge.
- Gao, J. (2015). Performance measurement and management in the public sector: Some lessons from research evidence. *Public Administration and Development*, 35, 86–96.
- Hofstede, G. (2001). *Culture's consequences: comparing values, behaviours, institutions and organisations across nations*. London: Sage.
- Halligan, J. (2007). Reintegrating government in third generation reforms of Australia and New Zealand. *Public Policy and Administration*, 22(2), 217–238.
- Improvement and Development Agency (IDeA). (2004). *A manager's guide to performance management*. London: IDeA.
- Kanter, R. M., & Summers, D. V. (1987). in Powell, W. W. (ed.), *The non-profit sector: A research handbook*. New Haven, CT: Yale University Press.
- Mackie, R. (2005). *The new public management of Scotland: Local government and the National Health Service*. Edinburgh: W. Green Thomson.
- Mackie, R. (2008). *Organisational performance management in a government context: A literature review*. Edinburgh: Scottish Government Social Research. www.scotland.gov.uk/Resource/Doc/236340/0064768.pdf.
- Mackie, R. (2013). *Managing Scotland's public services*. Edinburgh: W. Green Thomson Reuters.

- Organisation for Economic Co-operation and Development (OECD). (1997). *In search of results: Performance management practices*. Paris: OECD.
- Organisation for Economic Co-operation and Development (OECD). (2007). *Performance budgeting in OECD countries*. Paris: OECD.
- Organisation for Economic Co-operation and Development (OECD). (2015). *National Performance Frameworks and Key National Indicators Across OECD member countries*. Paris: OECD.
- Pollitt, C. (2001). Is the Emperor in his underwear? An analysis of the impacts of public management reform. *Public Management*, 2(2), 181–199.
- Pollitt, C. (2006). Performance Information for Democracy: The missing link. *Evaluation*, Vol. 12(1), pp. 38–55.
- Saunders, M., Lewis, P., & Thornhill, A. (2012). *Research methods for business students* (5th ed.). Harlow: Pearson Education.
- Scottish Government. (2007). *The government economic strategy*. Edinburgh: The Scottish Government.
- Scottish Government. (2008). *Scotland Performs launch*. Edinburgh: The Scottish Government.
- Scottish Government: Single Outcome Agreements (SOAs). (2009). *Governance and accountability for SOAs*. Edinburgh: The Scottish Government.
- Scottish Government. (2011). *Renewing Scotland's public services*. Edinburgh: The Scottish Government.
- Scottish Government. (2016a). *The programme for government 2016*. The Scottish Government: Edinburgh.
- Scottish Government. (2016b). *The Draft Budget 2017–18*. Edinburgh: The Scottish Government.
- Scottish Government (2016c). *The National Performance Framework*. Edinburgh: The Scottish Government
- Smith, P. (1995). On the unintended consequences of published performance data in the public sector. *International Journal of Public Administration*, 18(2/3), 277–310.
- Treasury Board of Canada Secretariat. (2003). *Management accountability framework*. Ottawa: Treasury Board of Canada.
- United States Government Accountability Office. (2017). *Performance and accountability report: Fiscal year 2007*. Washington: United States Government Accountability Office.
- World Bank. (2007). *How to build M&E systems to support better government*. Washington, D.C.: The World Bank.
- World Bank. (2015). *Worldwide governance indicators*. The World Bank: Washington, D.C.

Chapter 6

Governance Structures and the “(De) Politicization” of Performance Measures

Henrik P. Minassians and Ravi K. Roy

Abstract The nature of the politico-administrative structure at the county level shapes the way performance measures are used within them. In this chapter, we show how differences between two counties regarding the degree to which performance measures are linked to the overall strategic planning goals of the County as whole reflect differences in the politico-administrative structure of the counties themselves.

Keywords Performance measures · Governance structure · Scale of politico-administrative systems

6.1 Introduction

Democratic governments are supposed to implement policies that reflect the collective preferences of their citizens. Consequently, citizens expect government officials to use their taxes effectively to achieve common public goals. But what does this look like in practice? Public agencies in the USA have been emphasizing the use of both strategic planning and performance measurement initiatives to help justify their existence and the services they provide to their constituents. Performance measures are widely used to evaluate the results of government action in public, private, and nonprofit organizations to ensure the continued funding of particular programs.

H.P. Minassians (✉)

Urban Studies and Planning, Institute for Social and Behavioral Sciences, California State University Northridge, 18111 Nordhoff Street, Northridge, CA 91330-8259, USA
e-mail: Henrik.minassians@csun.edu

R.K. Roy

W. Edwards Deming Incubator for Public Affairs (WEDIPA), Southern State Utah University, 351 W. University Blvd., Cedar City, UT 84720, USA
e-mail: royr@suu.edu

Performance management focuses on improving the parts that operate within a larger organizational system. It came to be widely believed that organizational performance could be improved by disaggregating various functions into manageable parts. Consequently, public officials began decentralizing power and accountability within large public agencies. Under this new “leaner” and “meaner” organizational model, specific tasks and measurable goals could be developed and rigorously assessed through formal quantitative performance measures. Some of the most widely used performance evaluation methods of this sort include cost-benefit and cost-effectiveness analysis. Today, “performance measures, which encompass a variety of employee, customer, and other perspectives, are critical to management of the state’s activities” (Moynihan 2005, p. 36).

The current practice of measuring performance of public goals is far from a clear and straightforward process. The processes employed in assessing the effectiveness of public agencies can be severely complicated by the fact that many governments are not directly involved in the actual delivery of many goods and services. In the era of network governance, public agencies rely increasingly on a combination of other public, private, and nonprofit organizations to carry out their functions and mandates. Consequently, the success and failure of public agencies depends upon the operations and activities of a complex mix of governmental and nongovernmental organizations and their various internal organizational structures and cultures. Moreover, local public agencies operate in political, social, and economic environments where local legislative entities, along with state and federal governments, are continually shifting their priorities, funding, and goals. Formulas underlying current performance measures used to evaluate agency success often do not take these dynamics into account. When these vital dynamics are missing, we tend to get a distorted picture of agency performance.

The questions raised in this study are as follows: (1) What role does the politico-administrative structure of local county governments play in the design of performance measures? and (2) how do elected officials use performance measures in their decision-making processes? In addressing these fundamental questions, we will examine how performance measures are used in two counties in Southern California. Before we do so, let us first examine the literature on performance measures as a concept as well as an evaluation tool that is widely used across local state and federal public agencies in the contemporary era.

6.2 The Rise of Performance Measures: A Brief Overview

In 1993, the Government Performance and Results Act (GPRA) was enacted at the federal level with the aim of increasing accountability and improving the management of taxpayers’ dollars. For much of the nineteenth and twentieth centuries, concepts of accountability and performance in the public sector mainly focused on the amount of money public officials spent each year and how it was spent (De Lancer Julnes 2006). Hardly any focus was placed on how well the money was

spent. With the emergence of GPRA, however, the focus began shifting from how much governments spent and where toward measuring results. Newcomer (2007) found that 24% of US government agencies used performance information in the calculus of their funding decisions. This is true for the departments in the Southern California where the issue of accountability was placed on the agenda of the Board of Supervisors and various department heads.

The issue of “accountability” related to agency performance first appeared on the agendas of both of the two county governments under study in the late 1990s and early 2000s. Prior to that time, the issue of accountability was not prevalently discussed in either county. In the case of Los Angeles County, the US Department of Agriculture (USDA) gave the issue salience by placing it on the formal County Board agenda. Publicly released performance reports then followed that focused on improving services performed by the Department of Public Social Services (DPSS). Employing the new performance review process, county auditors discovered a high error rate in the way food stamps were being administered by the agency. From that point on, Los Angeles County DPSS would emphasize the use of performance measurements to ensure agency effectiveness. Adding performance measures requirement as a remedy for challenges that the County of Los Angeles was encountering, represents the real challenges and relations between politico-administrative scale and interrelations.

The USDA and the state of California began looking at specific programmatic performance issues within LA County Departments. LA County’s Board responded with a two-pronged approach to address these concerns. The first involved a broad strategy that focused on the design of the strategic plan of the county as a whole, while the other involved a more specific focus on the performance of the county, thirty-eight individual county departments. As a part of the county’s strategic plan, the Chief Executive Officer for Los Angeles County outlined a transformational vision that was directed at changing agency behavior. The new strategic plan outlined core goals for all thirty-eight county departments. Interestingly, performance measures were only sporadically used in budget documents or by elected officials. A former Los Angeles County social services department chief characterized the application of these performance measures in this way:

During my tenure as a County department head, there was never any consistent, uniform standardization of performance measures initiated by the Board (Bryce Yokomizo 01/30/2016).

6.3 Features of Politico-Administrative Systems and the Operations of Public Bureaucracies

Policy makers operating at various levels of government continue to be heavily involved in shaping the design of performance measures as well as how they are used. O’Toole and Meier (2014) assert that the relationship between public

management and program performance is shaped by four key variables: hierarchy, stability, network, and management. In analyzing the influence of these variables, they focus on the role played by the public manager within contemporary governance arrangements as well as the actions and objectives that are adopted by specific public organizations. As we shall see, these dynamics were heavily present in the design and adoption of performance measures that were undertaken by two counties in Southern California in the early 2000s. The study of public administration and public management has long been concerned with how the interaction between politico-administrative systems shape the decisions and activities of public bureaucracies. Typically, the literature in these related fields focus on structural, cultural, and functional variables (Pollitt and Bouckaert 2004).

Pollitt and Bouckaert (2004) examine how 5 politico-administrative features influence the design and implementation of performance measures in particular. These include the following: (1) *state structure*—which takes into account overlying city, county, and state jurisdictions; (2) the relationship between executive vs. legislative power as outlined within the state constitution.; (3) *the relationship between political appointees and elected officials*; (4) *the dominant administrative culture*; and (5) *the degree of diversity that exists within the organizational channels through which reforms emerge*.

Although accountability regimes may vary across different politico-administrative systems, performance measures can be an empowering tool that citizens can use to hold their public authorities accountable (Peters and Savoie 2000). This is particularly true at local and regional levels where governments are especially sensitive to citizens' demands and expectations. Others assert that citizens are "important players in shaping the quality and responsiveness of government programs in their community" (Epstein et al. 2000; Melkers and Willoughby 2005). As Hajer and Wagenaar (2003) suggest elected officials are often loathe to include citizens in the design and implementation processes related to governance and management performance. Dekker and Hansen (2004) disagree with this inference based on the assumption that measurement equates with effective use. Pollitt (2006) concludes that measuring practices have become universal, but politicians do not take interest in them except in cases of disasters or scandals.

The performance of governing networks, unlike single agencies or departments, now essential in the delivery of many public goods and services, is difficult to assess. That said, the collective role and effectiveness (or lack thereof) of multiple agencies and departments in providing these goods and services cannot be ignored. Bardach (1998) suggests that effective leadership involves promoting greater interagency collaboration. This is especially critical in environments characterized by high levels of organizational interdependence.

But managing networks (and hence assessing their effectiveness) is complicated given the fact that that individual public organizations often operate with high levels of autonomy (Provan et al. 2007). Moreover, Weiss (1998) argues that rigid limits imposed by laws, traditions, procedures, norms, and habits, that are characteristic of traditional organizational environments, are factors that can discourage organizations from collaborating in the design and implementation of performance

measures. Minassians (2015) argues, however, that hybridized environments, which involve complex organizational linkages between public, private, and nonprofit players, characteristically exhibit high levels of collaboration in processes related to the design and implementation of performance measures. O’Toole (1997) maintains that networks need to be afforded serious attention given the challenges that managers face when crafting decisions in the modern age.

Askim (2007) suggests that the role that politicians play is critical as they can determine whether and where performance management measures are used. Unfortunately, this feature is often overlooked because, as Moynihan reveals, the designers of performance management systems commonly assume that “performance information will automatically become a factor in existing decision processes” (Moynihan 2005, p. 211). In addition, the literature on public administration and management often overlooks the role that administrative scale plays in the use of performance measures.

6.4 Methodology and the Case Study

We focus on the role that elected officials in relation to the size and scale that politico-administrative structures play in influencing both the design and use of performance measures for assessing outputs and outcomes at the county level. In so doing, we will compare two disproportionately sized counties in Southern California to explore differences in their approaches to the design and use of performance measures. Our research design relies on a “within-case analysis” framework. In our comparative analysis, we will discuss the roles that the Board of Supervisors (elected body) vis-à-vis the role that departments within these counties have played in the design and implementation of performance measures.

The data used in this study is derived from the recent public documents and information available on the Los Angeles County Web site entitled *Program Summary and Performance Measures* (Los Angeles County 2012, <http://ceo.lacounty.gov/pdf/11-12/Program%20Summary.PDF>) as well as the Ventura County Web site (http://vcportal.ventura.org/CEO/docs/publications/Strategic_Plan_091311-1.c.pdf). From this data, we were able to analyze the format and content. In all, there are 1136 vs. 116 indicators of performance for Los Angeles County and Ventura County, respectively. We then identified each indicator according to whether they represented efficiency or effectiveness concerns of the organizations and whether they are interlinked allowing greater departmental collaboration.

The level of interdependence that exists among relevant stakeholders was also assessed (O’Leary and Bingham 2009). In addition, we conducted one-on-one interviews in order to identify the role that elected officials (the Board of Supervisors) play in shaping the purpose and use of performance measures in annual decision making. Four interviews were conducted either in-person or by

phone using “open-ended snowball sampling methods.” Those interviewed possess in-depth knowledge into why and how the performance measures work for their respective counties. These responses were then summarized and weaved into a narrative that is shared in this chapter.

6.5 The Different Politico-Administrative Structures of Two Counties and Their Distinct Approaches to the Design and Use of Performance Measures

Differences in the degree to which performance measures are linked to the overall success of strategic goals reflect differences in their respective politico-administrative structures. In the state of California, the responsibilities and mandates of county governments are outlined in the state constitution and the California government code. The county is the largest political subdivision of the state. The state legislature has the power to outline mandates that counties must follow as well as rescind powers and responsibilities that have been assumed by counties and their related departments.

Broadly speaking, there are two types of counties. The first is known as *general law* counties, which follow state laws outlining specified duties that must be carried out by county elected officials. The second is known as *charter* counties, which have a limited degree of “home rule” authority. As such, they have discretionary powers over certain elections, compensation terms, removal, and salary of the governing Board. Additionally, they oversee the election or appointment of county officers and consolidation and segregation of county offices. A charter, however, does not give county officials additional authority over local regulations, revenue-raising abilities, budgetary decisions, or intergovernmental relations. A county may adopt, amend, or repeal a charter with a majority vote. A new charter may be adopted, amended, or repealed by the Board of Supervisors through a charter commission or an initiative petition. Currently, there are 44 general law counties and 14 charter counties operating within the state of California.

Los Angeles County is a charter county where the Board of Supervisors exercises the role of an executive, legislative, and quasi-judicial authority. The Board can appoint either a Chief Administrative Officer or Chief Executive Officer to oversee the daily functions of each department (with the exception of elected members such as the Sheriff’s Office, the district attorney’s office, and the courts). The Chief Executive Officer carries more autonomy relative to the Chief Administrative Officer. Los Angeles County’s Board wants to maintain control and trusts less the directors’ ability to carry out their vision and goals. Also, there is less agreement on the goals due to the diversity of needs across five districts in a large geographical area.

By way of comparison, the Ventura County Board works collaboratively with the CEO and the department heads (Paul Derse 2015). According to the Executive

Director of California State Association of Counties Institute, the difference between Los Angeles County and Ventura County Boards is in large part due to their ability to collaboratively work with each other and allow directors and executive officers to conduct their job in a trusting environment. Ventura has a mature Board with a long history of a strong Chief Executive Officer and an effective understanding of the role of the Board and staff. “Not so much in Los Angeles where no one is sure whether the Board trusts itself or its CAO/CEO, thus it is more focused on directing senior staff then setting strategic direction” (Interview Bill Chiat 2016). This could be an explanation why the Board deputy asserted that the Board does not utilize performance measures systematically for strategic decision making.

6.6 The Nature, Purpose, and Use of Performance Measures: Strategic Goals and Indicator Comparison

The role of the departments in the design and implementation of strategic plans and how performance measures are used should be an important variable. Los Angeles County has devised 1,136 performance measures across 38 departments. The departments devised these performance measures, which focus on departmental outputs and outcomes. The missing link in the performance chain occurs when these indicators of performance are not directly linked with the strategic plan and goals of the County (see Table 6.1). The first column represents overall strategic goal, column two represents the county department out of 38 departments that actually use and comply with this strategic goal, and column three represents number of indicators that a particular department from column two uses meeting this particular strategic plan out of overall number of indicators that particular department utilizes. One explanation is that these indicators were devised in the absence of sufficient collaborative and interorganizational interaction. In the case of Los Angeles County, the link between performance measures and overall strategic goals remains weak.

By way of contrast, Ventura County uses fewer, but common core performance (116 in total) indicators of performance, which focus on the overall performance of the County government rather than a singular organization (see Table 6.2). These core indicators reflect broader strategic goals related to such things as public safety as a whole. Consequently, when attempting to measure the effectiveness of an early crime intervention program, Ventura County’s performance assessments tend to focus in the collaborative efforts of multiple departments and programs.

A close examination of six Los Angeles County Department’s performance indicators in relation to the overall strategic plan of the County shows that under Goal 1: “Strategic Initiative 5: Legacy System Replacement” there are no indicators assigned or data collected in this area. This could be concerning since four departments of Social Services, Mental Health, Children and Family Services, and the Sheriffs’ Department need to collaboratively share data in order to protect

Table 6.1 Los Angeles County strategic plan, goals, and performance measures

Los Angeles County strategic plan (2014)	Departments	Indicators per department
Goal 1: Operational effectiveness/fiscal sustainability: Strategic Initiative 1: Sound Fiscal Management/Capital Investments	Sheriffs Department Department of Public Social Department of Mental Health Department of Children and Family Services	5 indicators out of 140 6 indicators out of 64 0 indicators out of 38 5 indicators out of 64
Strategic Initiative 2: Targeted Risk Management	Department of Children and Family Services	12 indicators out of 64
Strategic Initiative 3: Countywide Contracting Improvement Initiative	Sheriffs Department Department of Public Social Department of Mental Health Department of Children and Family Services	3 indicators out of 140 7 indicators out of 64 1 indicator out of 38 8 indicators out of 64
Strategic Initiative 4: Innovative Technology Application	Department of Children and Family Services Department of Mental Health	3 indicators out of 64 6 indicator out of 38
Strategic Initiative 5: Legacy System Replacement	No Department uses indicators toward this strategic goal	
Goal 2: Community support and responsiveness: Strategic Initiative 2: Job Creation Efforts	Sheriffs	7 indicators out of 140
Strategic Initiative 3: Emergency Preparedness Expansion	No Department uses indicators toward this strategic goal	
Strategic Initiative 4: Healthy Neighborhood Projects	Department of Public Social Sheriffs Department of Children and Family Services Department of Mental Health	59 indicators out of 64 16 indicators out of 140 32 indicators out of 64 24 indicator out of 38
Strategic Initiative 5: Environmentally Sustainable Practices	No department uses indicators toward this strategic goal	
Goal 3: Integrated services delivery Strategic Initiative 1: Launch of Healthcare Reform	No department uses indicators toward this strategic goal	

(continued)

Table 6.1 (continued)

Los Angeles County strategic plan (2014)	Departments	Indicators per department
Strategic Initiative 2: Strengthening and Integrating Youth Protection Programs	Department of Public Social Sheriffs	8 indicators out of 64 2 indicators out of 140
Strategic Initiative 3: Implementing Jail Reform	No department uses indicators toward this strategic goal	
Strategic Initiative 4: Refinement of AB 109 (Public Safety Realignment) Implementation	No department uses indicators toward this strategic goal	

Table 6.2 Ventura County strategic plan, goals, and performance measures

Ventura County strategic plan (2013–2017)	Common core indicators of performance across all departments
<p><i>Focus Area #1: Good government, financial stability</i></p> <p>Strategic Goal 1: be a peak performing organization that consistently demonstrates effective use of available resources to provide the highest possible service and public communications</p> <p>Strategic Goal 2: maintain financial policies that are responsible and transparent, while building the County’s long-term durable financial strength</p> <p>Strategic Goal 3: invest in initiatives and tools to effectively and efficiently utilize, manage, optimize, and protect County workforce, resources, and assets</p> <p>Strategic Goal 4: promote an environment of economic vitality to support, retain, and attract businesses and support workforce development, each of which are vital for a prosperous and sustainable community</p>	Total of 18 indicators
<p><i>Focus Area #2: County workforce</i></p> <p>Strategic Goal 1: attract, hire, develop, and retain an effective, diverse, professional, dedicated, and responsive team of employees</p> <p>Strategic Goal 2: empower employees at every level to provide county services with maximum effectiveness and efficiency</p> <p>Strategic Goal 3: develop employees to become leaders who promote ethics, innovation, service, accountability, and peak performance</p> <p>Strategic Goal 4: champion and invest in workplace policies, programs, and practices that promote the overall health and well-being of all County employees</p>	Total of 11 indicators

(continued)

Table 6.2 (continued)

Ventura County strategic plan (2013–2017)	Common core indicators of performance across all departments
<p><i>Focus Area #3: Environment, land use, and infrastructure</i></p> <p>Strategic Goal 1: work with the ten cities and other responsible agencies to develop and implement shared programs which ensure the preservation of our unincorporated communities, agricultural land and natural environment, adequate housing for all residents, and the county’s continued economic viability</p> <p>Strategic Goal 2: provide, operate, and maintain infrastructure, public facilities, and associated services that protect and enhance our community, environment, and economic well-being</p> <p>Strategic Goal 3: provide simple and seamless services to our customers so that the benefit of living and doing business in Ventura County are fully realized</p> <p>Strategic Goal 4: champion cost-effective energy reduction measures through independent efforts as well as through regional initiatives and private/public partnerships</p>	<p>Total of 28 indicators</p>
<p><i>Focus Area #4: Community well-being</i></p> <p>Strategic Goal 1: achieve the triple aim, by providing quality healthcare in a patient- centered, integrated, equitable and efficient manner, improving the health of Ventura County residents</p> <p>Strategic Goal 2: ensure that individuals and families are provided timely and efficient assistance to meet/sustain basic needs, and transition quickly into pathways of productivity and self-sufficiency</p> <p>Strategic Goal 3: ensure that all children at risk receive the best treatment services to achieve the greatest success</p> <p>Strategic Goal 4: promote and provide for the preservation of healthy and safe communities so that all children may grow and thrive</p>	<p>Total of 40 indicators</p>
<p><i>Focus Area #5: Public safety</i></p> <p>Strategic Goal 1: maintain high performing public safety services</p> <p>Strategic Goal 2: pursue successful early intervention strategies to reduce future public safety threats</p> <p>Strategic Goal 3: engage in evidence-based intervention and supervision approaches in dealing with realigned post release community supervision (PRCS) and mandatory supervision (MS) offenders</p>	<p>Total of 20 indicators</p>

juveniles in public services or foster children and foster families, clearly displaying the importance of replacement of legacy systems and the design of integration management systems. To add a layer of complexity to the decentralized and fragmented American political system, there are programs that school districts (such as Los Angeles Unified) implement which are not under any legal or administrative authority of Los Angeles County. Even in an integrated system, this remains a challenge. This reflects a deficiency of knowledge and the need for more research among the scholars of performance measures and management system.

The case of Ventura County is different since indicators reflect overall strategic goals of the county. For example, “Focus Area 1: Strategic Goal 1: Attract, hire, develop and retain an effective, diverse, professional, dedicated and responsive team of employees” is a measurable indicator across all departments and can be tracked, measured, and compared across departments and different units. Five areas developed by the County of Ventura allow different departments to contribute their share of responsibilities and activities to the overall performance of departments in relation to the strategic plan of the County. Centralization of common core indicators encourages more collaborative participation in achievement of overall goals. There are two main explanations for these differences between Los Angeles and Ventura Counties. First, the sheer size and scope of Los Angeles County and the enormous size of its 38 departments make interdepartmental collaboration and planning extremely difficult, thus forcing many departments to maintain a siloed operation and protective of their resources from elected officials, thus becoming more politicized. Meanwhile, Ventura County Board has reached greater maturity and works collaboratively with the CEO and the department heads in order to implement strategic goals designed by the county.

6.7 Conclusion

We examined the role of elected officials in the use of performance measures through a comparative analysis of two California counties. Differences in the structures of politico-administrative systems were an important variable in influencing how performance measures were connected to strategic planning goals and utilized within county departments. In the case of Ventura County, the Board of Supervisors and the Chief Executive Office established a comprehensive strategic plan that outlined broad goals for the county as whole. Departmental directors were then empowered to design specific performance benchmarks in ways that were formally connected to these broader strategic planning goals.

In the case of Los Angeles County, the design of performance measures was more decentralized at the departmental level without any formal linkages to the overall strategic plan and goals of the county. This reflects the highly fragmented and siloed organizational structure of the county and a lack of connectedness between the various 38 departments. These attributes have not been conducive to supporting a cooperative relationship between the Board (elected officials) and the

CEO and department heads (the appointed officials). There has been very little coherence and consistency in performance assessments across departmental agencies within the county. Consequently, the performance data that were generated by the departments were highly insular and not connected to the broader strategic goals of the county. These findings confirm Pollitt and Bouckaert's (2004) proposition that the relationship between political appointees and elected officials directly affects the design of performance measures and how they are used.

The sheer size and scope of Los Angeles County, and the enormous size of its 38 departments, makes interdepartmental collaboration and planning extremely difficult. Therefore, the use of performance measures by department heads tend to be almost exclusively focused on promoting internal efficiency of the departmental parts without connecting them with larger county-level strategic planning goals. More recently, LA County has begun redesigning its strategic goals. Not surprisingly, many of these reforms have been undertaken at the departmental level and are unlikely to percolate through the systematic design and implementation processes.

The politico-administrative system of Ventura County is structured very differently. The Ventura County Board works collaboratively with the CEO and the department heads (Paul Derse 2015). The Ventura County Board has cultivated deep relationships with the Chief Executive Office. Hence, this relationship is characterized by high levels of interpersonal trust, cooperation, and coordination. These factors are aided by the fact that Ventura County boasts a relatively coherent and well-coordinated administrative structure, which tend to promote greater cooperation among the various departments. In the case of Ventura County, cohesive design of performance measures and management system has helped them to identify common core indicators across departments or singular agencies. The performance of organizations is linked to larger county-level strategic goals.

In conclusion, New Public Management (NPM) and public governance convergence theories can explain the movement of various public entities toward the adoption of performance measures. That said, there is a paucity of research on how politico-administrative systems impact the design of performance measures. More specifically, there is very little academic work on how government and governance type, scale of these entities, affect the design and adaptation of performance measures.

Further research should investigate the effect of task characteristics on county departments' design influence, a question beyond the scope of this article. Secondly, engaging departments in the overall design of common core indicators of performance may prevent adverse behavior of administrative actors in the design of insular performance indicators as well as during implementation phase (deLeon and deLeon 2002). In situations where the departments tend to design indicators in an insular setting, problems of disconnection between different levels are less likely to occur. Finally, cooperation and collaboration in all stages of strategic plan design and linkage of performance measures to the overall success of the County mission can increase trust levels between elected officials and department heads (Rommel and Christiaen 2009).

References

- Askim, J. (2007). How do politicians use performance information? An analysis of the Norwegian local government experience. *International Review of Administrative Sciences*, 73(3), 453–472.
- Bardach, E. (1998). *Managerial craftsmanship: Getting agencies to work together*. Washington, DC: Brookings.
- Bill Chiat. (2016). Personal Interview.
- Bryce Yokomizo. (2016). Personal Interview.
- County of Ventura, Strategic Plan 2011–2016 (Data File). http://vcportal.ventura.org/CEO/docs/publications/Strategic_Plan_091311-1.c.pdf.
- De Lancer Julnes, P. (2006). Performance measurement: An effective tool for government accountability. *Evaluation*, 12(2), 219–235.
- Dekker, S., & Hansén, D. (2004). Learning under pressure: The effect of politicization on organizational Learning in public bureaucracies. *Journal of Public Administration Research and Theory*, 14(2), 211–230.
- deLeon, P., & deLeon, L. (2002). What ever happened to policy implementation? An alternative approach. *Journal of Public Administration Research and Theory*, 12, 467–492.
- Epstein, P., L. Wray, Marshall, M. & Grifel, S. (2000) “Engaging citizens in achieving results that matter: A model for effective twenty first century governance”. Paper presented at the center for accountability and performance’s symposium on leadership of results-oriented management in government, Washington, DC, 11 February.
- Hajer, M. A., & Wagennar, H. (2003). *Deliberative policy analysis: Understanding governance in the network society*. London: Cambridge University Press.
- Los Angeles County, Chief Executive Office. (2012). 2011–2012 recommended budget (Data File). <http://ceo.lacounty.gov/pdf/11-12/Program%20Summary.PDF>.
- Melkers, J., & Willoughby, K. (2005). Models of performance-measurement use in local government: understanding budgeting, communication, and lasting effects. *Public Administration Review*, 65(2), 180–190.
- Minassians, H. P. (2015). Network governance and performance measures: challenges in collaborative design of hybridized environments. *International Review of Public Administration*, 20(4), 335–352.
- Moynihan, D. P. (2005). Why and how do state governments adopt and implement “Managing for Results” reforms? *Journal of Public Administration Research and Theory*, 15(2), 219–243.
- Newcomer, K. (2007). How does program performance affect program management in the federal government? *Public Performance and Management Review*, 30(3), 332–350.
- O’Leary, R., & Bingham, L. B. (2009). *The collaborative public manager: Bew ideas for the twenty-first century*. Washington, D.C.: Georgetown University Press.
- O’Toole, L. J. (1997). Treating networks seriously: Practical and research-based agendas in public administration. *Public Administration Review*, 57, 45–52.
- O’Toole, L. J., & Meier, K. J. (2014). Management, context, and performance. General theory. *Journal of Public Administration Research and Theory*, 26(4), 2–20.
- Paul Derse. (2015). Personal Interview.
- Peters, G. B., & Savoie, D. J. (2000). *Governance in the twenty first century: Revitalizing the public Service*. Montreal: McGill-Queen’s University Press.
- Pollitt, C. (2006). Performance management in practice: A comparative study of executive agencies. *Journal of Public Administration Research and Theory*, 16(1), 25–44.
- Pollitt, C., & Bouckaert, G. (2004). *Public management reform: A comparative analysis*. London: Oxford Press.

- Provan, K., Fish, A., & Sydow, J. (2007). Interorganizational networks at the network level: A review of the empirical literature on whole networks. *Journal of Management*, *33*(3), 479–516.
- Rommel, J., & Christiaens, J. (2009). Steering from ministers and departments: Coping strategies of agencies in Flanders. *Public Management Review*, *11*, 79–100.
- Weiss, C. H. (1998). Have we learned anything about the use of evaluation? *American Journal of Evaluation*, *19*(1), 21–33.

Chapter 7

Italian Public Administration Reform: What are the Limits of Financial Performance Measures?

Paolo Ricci and Renato Civitillo

Abstract The Italian public sector reforms in recent years have demonstrated an over-reliance on accounting-based financial measurements which has essentially created a sort of ‘hierarchization’ of performance. This chapter aims to demonstrate whether and how this predominance leads to negative consequences in the evaluation (and management) of public sector organizations: First, because in definitive governments, performance should be assessed coupling financial parameters with non-financial measures and qualitative judgements (Jones and Pendlebury in *Public Sector Accounting*, 6th ed., Pearson Prentice Hall, London, 2010); second, for the lack of a systemic approach, financial performance should not be the ultimate objective of public management but instead an instrument to evaluate the financial comparability of various priorities to pursue (public value, social, environmental, etc.) (Esposito and Ricci in *Public Money Manage* 35(3):227–231, 2015).

Keywords Non-financial performance · Hierarchy · Public value

7.1 Introduction

The role of the public administration in Italy, as in other European countries, has changed profoundly over time, with an evolution of its roles and functions that has resulted in a significant increase in the areas that fall within its realm of action. This has led to the emergence of critical issues, namely (Hughes 2003):

P. Ricci (✉) · R. Civitillo
Department of Law, Economics, Management and Quantitative Methods-DEMM,
Sannio University, Via Delle Puglie, 82, 82100 Benevento (BN), Italy
e-mail: ricci@unisannio.it

R. Civitillo
e-mail: civitillo@unisannio.it

- the excessive use of resources has had a negative and widespread impact;
- the excessive reach of public intervention, with involvement in areas that are too far from its traditional role;
- high levels of inefficiency in the quality of services offered.

New Public Management and the other theoretical movements which have revolutionized Italian public administration since the 1980s (Public Governance, Public Performance Management and New Public Governance) have led to a paradigm shift (Barzelay 1992; Behn 2001), which by relying on neoclassic economic theories, from Public Choice Theory (Stigler 1971) and Principal Agent Theory, seek to reach the highest levels of efficiency, effectiveness, and productivity within the public sector through the introduction of an entrepreneurial culture, methods, and techniques (Hood 1991).

In fact, the traditional conceptual pattern according to which the respect of norms is sufficient to automatically reach an optimal balance in government action has become obsolete in light of the administrative approach, which instead posits effectiveness (the ability to satisfy community needs), efficiency (the ability to reach objectives, using the least amount of resources), and cost-effectiveness (the ability to maintain the correct balance between the resources used and the benefits obtained for the community in the long term) as the basis of a properly functioning public organization.

The need to measure performance has inspired key reforms in the public sector, at a national and international level (OECD 2005; Pollitt and Bouckaert 2004): The level of performance of government action represents a fundamental element for the evaluation of the correct and rational use of public funds and thus to ensure the adequate level of transparency, efficiency, and effectiveness of the choices made for the good of the community (Hood 1991; Pollitt and Bouckaert 2004).

In this context, the problem of limited resources available and the financial instability of the new organizational and management models of the public administration have made it so that the interest in financial performance has become so important that a disproportionate amount of attention is given to the methodologies and instruments for its measurement.

For the purposes of this work, 'financial performance' is defined as the achievement of economic and financial objectives, measured according to the methodology and techniques of 'financial accounting theory.' The main accounting practices used by public sector organizations are (Jones and Pendlebury 2000):

- budgetary accounting;
- cash accounting;
- accrual accounting;
- commitment accounting;
- fund accounting.

The results of these accounting practices are indicators used to determine financial performance, for example, a school's budget for a year, the average cost of a doctor's visit, and the cost of obtaining court judgment in a legal dispute.

Financial performance has therefore gained more and more relevance in policy and in legislative reforms—creating a sort of ‘hierarchy’—and prevailing over other dimensions.

By ‘hierarchy,’ we mean a system in which things are arranged according to different levels of importance, from highest to lowest (Cambridge Dictionary 2016). With reference to public administration,

hierarchy establishes the democratic current that runs throughout contemporary systems of public governance and administration, linking the various actors, organizations, and institutions that make up the core features of democratic systems of governance (Bovens et al. 2014, p. 405).

In this respect, however, some authors also believe that

non-financial inputs, outputs and outcomes of government services are best thought of as being hierarchical (Jones and Pendlebury 2010, p. 21).

The causes of ‘hierarchization’ can be summarized in the following way:

- (a) a push toward the simplification of checks and measurements in order to achieve the objectives of each public administration;
- (b) the absence of a systemic view of every single public organization. In Italy, for example, this view was introduced more clearly only recently through the adoption of accounting harmonization (art. 9 Legislative Decree n. 118/2011);
- (c) the prevailing financial culture in the EU Stability and Growth Pact (SGP).

In light of this, the present work, which is based on an intuitive–deductive approach, has the following objectives:

- 1) highlight how it is possible to identify a hierarchy among the elements of performance, with an over-reliance on financial performance;
- 2) demonstrate how the predominance of financial performance could lead to negative consequences in the evaluation (and management) of the public administration, undermining the necessary systemic and collective vision; when measuring the performance of public sector organizations, we should distinguish between distinct elements of performance: inputs, outputs, and outcomes. In light of this, it would be useful to use not only financial parameters but also non-financial measures and qualitative judgements (Jones and Pendlebury 2010);
- 3) underline the role of financial performance as a means to evaluate various priorities (public value, social, environmental, etc.) and not as an end goal of public administration (Adams et al. 2014; Esposito et al. 2015).

7.2 Measuring Performance

Defining performance is extremely complex (Ridley and Simon 1943; Lapsley and Mitchell 1996; Atkinson et al. 1997; Streib and Poister 1999; Kloot and Martin 2000; Halachmi 2005; Monteduro and Hinna 2007). While, in general terms, it can

be defined as the ability to achieve a result (Bovaird 1996), as in a response to a need (Liguori et al. 2012), it can also be understood in very different terms based on the theoretical approach and the chosen ends. For this work, it is helpful to define performance as the evaluation, based on the criteria of efficiency, effectiveness, or a social nature, of an organization's ability to meet the expectations of those who are part of it. Performance is about intentional behavior, which can be individual or organizational (Van Dooren et al. 2010) and which implies a certain standard for quality:

(a) the quality of the actions being performed,

or

(b) the quality of what has been achieved because of those actions.

Performance has to do with important and specific aspects of governing such as the provision of services, in definitive form, free at the point of delivery to specific individuals or groups of individuals (Jones and Pendlebury 2010).

From the definitions above, fundamental characteristics emerge very clearly (Guthrie and English 1997; Van Dooren et al. 2010):

- the concept of subjectivity,
- a multidimensionality within the concept of performance,
- the quality of actions and results achieved.

The subjectivity has to do with the fact that every level of performance depends largely on a combination of expected results, actors involved, policies, programs, and services offered, which are tied to previously determined needs (De Bruijn 2007; Thomas 2007). Their measurement and their evaluation are strongly conditioned by information needs and the characteristics of the subject or subjects involved. The logical basis for the measurement of performance consists in the proper identification of key factors and the subsequent creation of parameters using these key factors (Kloot and Martin 2000). Subjective expectations are characterized by a certain degree of ambiguity that could make them more or less undetermined, clear, and constant in time. In this sense, another point to take into consideration is possible behavior changes caused by the measurement (Hatry 2002; Thiel and Van Leeuw 2002; Van Dooren 2006; De Bruijn 2007). This represents an additional element of complexity within the concept of performance that goes alongside the subjectivity mentioned above.

The multidimensionality of performance in the public sector (Moore 1995; Guthrie and English 1997; Bouckaert and Halligan 2008) refers to the need for a systemic methodology through the integration of economic variables with technical indicators (Epstein and Birchard 2000), strategic and operative needs (Kaplan and Norton 2001), as well as internal and external perspectives. More specifically, the multidimensionality can be defined based on its content (width) as well as its application in time (depth) (Bouckaert and Halligan 2008).

According to the ‘quality’ characteristic, we can distinguish two perspectives (Van Dooren et al. 2010): firstly, the tasks being carried out by the performing agent. The municipal waste service, a vaccination campaign, a surgical procedure, and a university lecture are all (very broad) examples of performance by government actors. In this sense, performance is conceptualized as ‘competence’ or ‘capacity.’ However, each performance may have a high or low level of quality; secondly, because ‘competence’ and ‘quality outputs’ are directly proportional (Dubnick 2005), when performance is about the quality of the achievements and not as much about the quality of the actions, performance equals results (Van Dooren et al. 2010).

Notwithstanding the critical elements illustrated above, it is necessary to underline how the measurement of performance, which consists in the process of determining and assigning to it a quantitative value, represents the prerequisite for any type of evaluation or judgment that has to do with the services rendered by the public administration. This seemingly simplistic consideration highlights an obvious difficulty in the measuring procedures used for evaluating public organizations, which in turn are characterized by significant qualitative elements due the social nature of their aims that are, by their very nature, difficult to measure. In fact, because of the multitude of interests involved in the correct functioning of a given public administration, the evaluation of the results achieved is so crucial that it should involve every actor that is potentially in contact with it. To ensure accountability (De Bruijn 2007), the systems used to measure and evaluate performance must be conceptually, theoretically, and empirically coherent (Del Bene 2014). From this stems, the use of performance logic and related measuring tools for public organizations—a logic which should avoid facing the risk of measuring too much or measuring only what is ‘measurable.’ For this reason, a number of theoretical approaches for measuring public organization performance have emerged since the 1990s. These approaches provide their own vision of evaluation of performance, but aside from their relative specificity, they all have the same objective: to introduce performance management systems which go beyond the traditional system of control based on compliance or so-called conformance. Conformance is the sterile adherence to rules and procedures without an evaluation approach based on results. In sum, there has been a theoretical shift from a ‘culture of mere adherence’ to a ‘culture of performance’ (Monteduro and Hinna 2007; Borgonovi 2009). Therefore, we should distinguish between ‘adoption’ and ‘implementation’ of performance measurement systems (De Lancer and Holzer 2001). In this perspective, it is useful to point out that system dynamics can be used to enrich performance management in public sector organizations, as shown by some recent studies (Bianchi and Rivenbark 2014).

Even considering the subjective and multidimensional elements discussed above, what follows is that the process of evaluation implies the measurement of value generated. In the private sector, this is based on market mechanisms, whereas in the public sector, the rules are completely different (Jones and Pendlebury 2000; Borgonovi 2001).

It suffices to recall that all public administrations have the aim of furthering the common good and that in Italy, this is intimately linked to the principles of proper functioning contained in Article 97 of the Italian Constitution. This means that all activities must be inspired by the notions of efficiency, effectiveness, and economy— notions that fall clearly within the administrative doctrine and can be interpreted as the basis of legality in public administration (Borgonovi 2001). In other terms, legalitmplicitly connected to mere adherence to the rules but, rather it is the consequence of the concrete actions which respect the above-mentioned criteria of efficiency, effectiveness, and economy and thus the principle ‘well-functioning’ public administration contained in Article 97 of the constitution.

The absence of measurement and evaluation systems for the performance of the public administration has a negative impact on the processes for reform, which are made weaker as a result. It is for this reason that, for some time now, reforms have been put in place in several European countries to counter this absence. Examples include the UK, Germany (Neues Steuerungsmodell), and France (LOLF—Loi Organique relative aux Lois de Finances).

In this perspective:

Performance measures contain information that can be used not only to evaluate, but also to learn. Indeed, learning is more than evaluation. The objective of evaluation is to determine what is working and what isn’t. The objective of learning is to determine why (Behn 2003, p. 592).

Along these lines, Legislative Decree n. 150/2009 represents an attempt (although not a timely one compared to other countries) to shed light on the need for performance evaluation in the public sector, focusing on issues tied to the reliability of performance measurement (Del Bene 2014).

7.3 The Italian Approach to the Culture of Performance: Legislative Decrees N. 286/1999 and N. 150/2009

The introduction of performance measurement processes in the Italian public sector may represent one of the most important attempts to move beyond the traditional ‘adherence approach’ toward a true ‘culture of performance’ focused on results (Dunleavy and Hood 1994; Pollitt and Bouckaert 2004; Guthrie et al. 2005). One of the first comprehensive laws in Italy on the topic of performance measurement can be traced to Legislative Decree n. 286/1999 which sets up a system of evaluation and internal checks within the public administration with the aim of monitoring several aspects of the public management through a strategic control body, to which it assigned the task of evaluating the adequacy of the choices made in the implementation of plans, programs, and other policy instruments, in terms of the coherence between results obtained and initial objectives. This sort of strategic control process in public organizations can be interpreted as a level between the typical outcome of public action and the relationship between inputs and outputs

that characterize the managerial approach. In fact, the process of control is projected into a dimension that is not a purely normative or legal one, but one which checks the total impact of the actions of the public administration aimed at implementing the public policies for the good of the community.

The system of internal checks included in Legislative Decree n. 286/1999 has a rather fragmented structure and is especially lacking an adequate systemic approach. The main objectives of the law can be summarized as follows:

- to guarantee legitimacy, accuracy, and regularity of the administrative and accounting spheres;
- the optimization of the relationship between costs and results, including through feedback mechanisms;
- the evaluation of the performance and organizational abilities of management personnel;
- the verification of the adequacy of the strategic and policy choices and the adequate balance between results and objectives.

However, the implementation of the innovative aspects of Legislative Decree n. 286/1999 allowed for various critical points to emerge, especially due to a bureaucratic approach toward evaluation and control. The model, in fact, is based on a one-dimensional analysis of performance, with a top-down logic and evaluations of employees carried out by management. Various difficulties emerged with regard to the correct identification of parameters to measure results, with dire consequences on the effectiveness of the entire norm. Furthermore, in the Legislative Decree, even though a great deal of attention is given to the promotion of internal accountability (among political and administrative organs), very little consideration is given of outside accountability—with little transparency in accounting for results obtained to the citizens and community (Ricci 2016).

It is precisely because of the shortcomings listed above that Italian legislators decided to modify Legislative Decree n. 289/1999 several times and reached the conclusion that the problem with the law (and with Legislative Decree n. 29/1993 and the others written in the 1990s) was not its content but its effective implementation. For this reason, with the Legislative Decree n. 150 of October 27, 2009 ('Brunetta Decree'), Italian legislators tried to reorganize the norms in the area of the optimization of productivity, efficiency, and transparency of the public administration. In order to overcome the critical points of the previous laws, particularly the implementation issues of Legislative Decree n. 286/1999, the Brunetta Decree tried to find a solution to the cultural problem rather than to the technical one. In this sense, in addition to performance, to which particular attention is given in Legislative Decree n. 150/2009, other concepts are taken into consideration which are by no means new to the Italian legislative landscape, such as efficiency, effectiveness, economy, productivity, and transparency. However, the law introduced some significant conceptual innovations related to the general legal approach and the instruments that should be used. Despite these significant innovations, for the purposes of this work, the most important innovation of the Brunetta Decree is

the introduction of the necessary and fundamental systemic vision of the entire picture which was largely absent in the prior regulatory approaches. In this sense, the model of management set forth in the law is inspired by the objective of guaranteeing a model aimed toward results and a managerial approach to public administration, as highlighted in the leading international literature (OECD 1997, 2004, 2005; Bouckaert and Halligan 2008). The general goal of the norm was to improve the results of organizations and more specifically to:

- improve the quality of the services offered by public administrations;
- promote professional growth and development within public organizations.

These objectives are reached through an organic approach in which the single elements—such as the process of defining goals, deciding what resources to allocate, the creation of mechanisms and systems for rewarding achievement based on results achieved, and internal and external accounting practices—are organized in a systemic way and not as separate elements without any interdependent connection (Otley 1999). The systemic and organized approach adopted in Legislative Decree n. 150/2009 ensures that the single objectives of one phase do not outweigh the general mission of the entire organization (Riccaboni 1993). To this end, another important element that was introduced by the Brunetta Decree is the involvement of the entire organizational structure and the assignment of specific responsibilities to various subjects involved in the so-called ‘management cycle of performance.’ This concept originates from the ‘management plan for performance’ a three-year planning document in which objectives, indicators, and targets at the basis of measurement, evaluation, and accounting of performance are clearly stated. Despite the key role that the concept of performance plays, the norm does not focus on its measurement but rather gives greater weight to the definition of outputs to account for and to use as a basis for its evaluation (Borgonovi and Valotti 2009). Still, it is necessary to highlight the difficulty of measuring performance in the context of public administration. It is, in fact, difficult (and in some cases impossible) to identify quantitative criteria to evaluate essentially qualitative results. Furthermore, even when these criteria are predictable or identifiable, how they are characterized can have an impact on their ultimate relevance and effectiveness for interpretive uses. From this perspective, it is helpful to distinguish between a quantitative dimension of performance, which is characterized by indicators that allow for the measurement of specific management aspects, and a qualitative dimension which, by nature, allows for a non-quantitative evaluation of the actions of the organization.

In this sense, it is important to highlight that despite the ambitious goals of the reform, it fell short in various aspects and led to mixed results (CIVIT 2011, 2012). Several studies have pointed to limitations, particularly pertaining to performance, transparency, and quality (ANAC 2013).

In a study from 2012 (Galli 2012), the concepts of relevance and measurement of performance contained in the Brunetta reform were compared to those in place in 7 countries: UK, Canada, Australia, USA, France, Germany, and Finland. The

study found that none of the countries examined had accounting mechanisms in place that would meet the requirements of Legislative Decree n. 150/2009. This shows that the requirements of the reform were too ambitious, even when applied to countries in which performance management has been introduced for quite some time (Galli and Turrini 2013; Bigoni and Deidda Gagliardo 2013; Cuganesan et al. 2014).

Finally, the Brunetta reform represents a positive legislative innovation from a cultural point of view, but its impact on the Italian public administration has been quite limited. This has been the case particularly because of the financial limitations of the reform, which in turn have led to only a partial application of its provisions (Ricci and Serluca 2013).

7.4 The Performance Hierarchy: Does the Financial Dimension Have to be at the Top?

As mentioned previously, the last two decades have been characterized by various attempts by Italian legislators to develop models which answered to different theories and needs, each characterized by their own peculiarities. We can identify two different phases in this legislative process. The reforms of the 1990s, culminating in Legislative Decree n. 286/1999, were characterized by a push toward ‘tasks,’ typical of the above-mentioned approach based on abiding to norms and ‘accountability bureaucracy’ which focused on creating monitoring bodies and inspection-like checks, aimed toward ensuring that administrative acts fell in line with the applicable legal norms. Later, the need to measure performance focused exclusively on outputs (on the goods and services rendered), and it was carried out only by external bodies, laid out in the law itself (Monteduro 2010).

As shown previously, the evolution of the role of the public administration at the international and national level set the basis for a move past the traditional approach, and in favor of various attempts to introduce a true culture of performance, one focused on results rather than on mere adherence to norms. Legislative Decree n. 150/2009, which was characterized by the shift from New Public Management to Public Governance, is the norm which represents this transition. Here, the attempts at creating a management approach to public administration that were started in the 1990s are accompanied by a different approach, a multidimensional one that aims at external accountability in order to guarantee and favor transparent knowledge and understanding of the value created by public bodies (Moore 1995; Guthrie and English 1997; Kelly et al. 2002; Stoker 2006; Beck Jørgensen and Bozeman 2007; Bozeman 2007; O’Flynn 2007) and improving the satisfaction of user needs and use of resources (Holzer and Yang 2004). In this way, the importance of what is being measured (Berman 2002; Lemieux-Charles et al. 2003; Fryer et al. 2009; Van Dooren 2006) is connected with the outcomes of the government actions and the recipients of information are no longer only external

bodies but stakeholders (especially citizens) who become more involved in the entire management model (Moore 1995).

In this sense, many studies have shown that well-managed performance measurement systems are critical for accountability in public sector organizations (Rivenbark 2007; Aguinis et al. 2011; Bianchi and Rivenbark 2014).

However, despite the above-mentioned multidimensionality which characterized the latest norms in Italy, all legislative reforms have had a common denominator: the prevalence of the financial dimension of performance that emphasizes accounting-based measurement and thus is able to capture the ‘economic value’ generated by the public administration but not the public value of its actions, which should be the predominant characteristic and purpose of public administration (Cuganesan et al. 2014; Bracci et al. 2014). In Italy, but also in other countries, this prevalence of the quantitative–financial dimension has progressively become a predominance and has begun to influence, in an increasingly significant way, all the other dimensions as well, so much so that it has led to the hierarchization of different dimensions of performance, with the financial dimension at the very top. In fact, the rationale for this approach is nothing new (Drucker 1954, 1976) and reflects the view of several public sector reformers (Holmblad Brunsson 2002; Modell 2004).

In this sense, we can therefore refer to this as a ‘hierarchization of performance.’

Financial measurement systems are a general characteristic of all organizations and are presented differently based on different classifications and theories. For public organizations, especially in Italy, in the past, the topic of accounting-based ‘financial balance’ was often second to the concept of public finality and so unexpected, because it was often possible to incur public debt. Later, instead, this phenomenon became more strained, following a reduction of the resources available because of the limits imposed by macroeconomic relations. Therefore, it appears necessary to find a balance between achieving objectives in the interest of the public and the financial measurement of the same.

On the other hand, it is true that ignoring financial parameters and cost has many negative consequences, also because the cost of services is really important for performance measurement. This means that inputs, outputs, and outcomes must be judged together to lead to useful performance analysis:

In the short term, accounting might judge successful outputs, with unsuccessful outcomes but at low cost, favorably (number and classification of examination passes increase, but approval level fall and budgets are lowest compared with comparable services), but in the medium term might worry about the implications for future budgets if the low approval levels lead to radical changes (Jones and Pendlebury 2010, p. 21).

Therefore, while attention to these accounting-based elements is necessary, it must be aimed at identifying areas of waste, illicit activity, and introducing services and processes to award merit and professionalism. The achievement of these objectives can be measured through adequate and specific indicators of financial performance

(e.g., the reduction of teaching personnel salaries in schools and universities, reduction of treatments offered in hospitals or the costs of managing public transportation, etc.). These types of financial and economic indicators should guide the public administration, leading to an achievement of the objectives set forth in the planning process, but they raise the following questions (Borgonovi 2009): Is a balanced budget always in the interest of citizens? How can we judge the achievement of a balanced budget if this is due to the reduction in the quantity or quality of services offered?

These questions highlight the need for a systemic approach to performance measurement and adequate evaluation processes.

In support of this concept, Jones and Pendlebury (2010, p. 21), in their book ‘Public Sector Accounting,’ state:

‘In definitive governments, in which the services provided free at the point of delivery are financed by taxation, performance is assessed using financial and non-financial measures and qualitative judgements. In performance measurement, it is useful to think of the following distinct elements of performance:

- inputs, being resources consumed by the governments, measured primarily using costs but also non-financial measures—commonly, the number of employees
- outputs, being the services provided, measured primarily using non-financial measures
- outcomes, also being the services provided, but primarily using unmeasured, qualitative judgements, though when the judgements of outcomes are systematically gathered from service recipients, typically based on interviews or questionnaires, they can be measured and statistics of satisfaction produced.

Non-financial inputs, outputs and outcomes of government services are best thought of as being hierarchical. [...]. The lowest levels in the hierarchy, while they can be reliably measured, are furthest away from what the government services are ultimately trying to achieve; at the highest level they are what the services are ultimately trying to achieve, but cannot be measured.

None of the levels in the hierarchy of outputs and outcomes is the natural responsibility of accounting. In the provision of government services (as in non-profits), outputs and outcomes are matter for others—in this case, service professionals and politicians.’

From what we have just highlighted, accounting-based performance measurements are not very useful for our purposes: These are useful just for partial assessments, and they should be completed by qualitative judgments and, above all, with non-financial measurements.

Furthermore, in the public sector, the systematic approach highlighted above has an even greater importance: In this case, the evaluation of performance of a single unit must be considered in its totality (or again, in a systemic way) and it cannot disregard the results of other existing organizations. In other terms, if the improvement of the economic–financial balance of an administration is not in line with the similar results obtained by other organizations, the equilibrium is tarnished in the broader system of which it is a part, aside from the one tightly linked to the various entities taken into consideration individually.

7.5 For a Better Understanding: Five Short Examples

To better understand what we mean by ‘hierarchization’ of performance, and how it permeates the regulatory landscape, management models, and evaluation of the Italian public sector organizations, we offer some examples involving four relevant fields, concerning the evaluation of performance: municipal bankruptcy legislation, healthcare system, public transportation services, and international migration facilities. They are examples that try to clarify the potential dynamics that exist between financial and non-financial performance.

The Italian legal system (as in other countries) recognizes the possibility that municipalities and other local entities may encounter moments of financial crisis, of various levels of gravity. Aside from the procedures and instruments contemplated, for the purposes of this work, it is sufficient to highlight the parameters and criteria to determine the degree to which and whether there is a financial crisis: It is based on quantitative indicators that reveal exclusively financial results. On the other hand, no consideration is given to the evaluation of other aspects of performance such as the quality or the variety and abundance of services offered to citizens (Peck 2014).

A further example is the provision of healthcare services in Italy. In the so-called Health Pact 2009/2012, it is clear that the aim of improving the functioning of Italian national health system has been interpreted exclusively according to a financial perspective, with no consideration of the effective protection of health (Anderson and Frogner 2008). As Borroni and Compagni argued (2013, p. S35):

Attention appears to be focused on how to collect sufficient resources to sustain health care systems.

In fact, it is a financial planning document that excludes any assessment of the quality of health services provided to citizens. To understand this, it is sufficient to note that in the text of the provision, the word ‘patient’ is present only once, while the words ‘disease,’ ‘human person,’ ‘human resource,’ and ‘responsibility’ never appear. Further confirmation is given by the structure of the legislative provision. The basic elements include:

- the estimated budget of expenditure of the National and Regional Health Service;
- a system of indicators covering the average costs and standard costs of the services provided;
- the provision of a financial recovery plan (in case of budgetary imbalances).

The introduction of performance management tools is the purpose of the recent reforms that have also affected the system of Italian universities. In this direction, the new adoption of the accrual basis accounting is aimed at guaranteeing the highest levels of efficiency and effectiveness (Romano and Cirillo 2015). However, even in this case, to achieve these management objectives, attention is focused totally on financial elements (e.g., cash flow, economic balance, and standard

costs). In contrast, no attention is given to non-financial aspects such as public value or, especially, the third mission, which is:

a global trend where universities are collaborating with government, industry and civil society to advance the sustainable transformation of a specific geographical area or societal sub-system (Trencher et al. 2014, p. 151).

The end result, therefore, is represented again by the dominance of the financial dimension. It is given more importance than the other dimensions in the evaluation of the overall performance of the Italian universities.

The same occurs for the evaluation of the public transportation system. In this case as well, the aim sought by recent reforms is to guarantee an improvement of the economic conditions of the companies which manage the services—an objective which is reached by reducing the frequency or number of buses in circulation. It is evident, however, how the possible reduction of costs of the service can create various negative externalities, such as the worsening of the environmental conditions in the city (higher levels of pollution and related ailments). In this sense:

Urban transportation system is a complex system with multiple variables and nonlinear feedback loops and influenced by transportation, social, economical, and environmental factors (Wang et al. 2008, p. 83).

This means that financial improvement results in a worsening of the financial conditions of the healthcare system and therefore of the financial balance of the system as a whole.

The recent international flows of migrants and refugees, from Middle East and Africa to European countries, can also be viewed from the same perspective. The superficial management of the services and structures dedicated to controlling the influx of immigrants, which in turn were the result of the need to contain spending, could have social economic repercussions (Campesi 2011; Marchetti 2014), with an impact on the health of the migrants or on the crime rates in the affected areas, determining an increase in the costs the national healthcare system for the care of the migrants and for the police force needed to guarantee the safety of citizens.

7.6 Preliminary Conclusions

The considerations made thus far set up an evident challenge, which is also the solution to the critical points just discussed: the search for models that lead to effective systemic balance, both financial and non-financial nature.

This consideration requires a complete rethinking of the role of financial performance indicators in public sector organizations. In fact, the phenomenon of ‘hierarchization’ of performance of the Italian public administration illustrated above determined the inability to construct an effective multidimensional model for performance measurement. A multidimensional approach, in fact, is the only type able to lead to an efficient and incisive achievement of the integrated and systemic

objectives, precisely because it is only with the recognition that there are many dimensions to performance that a predominance of one dimension (and therefore limited) over the others can be avoided. It is especially problematic when it is the accounting-based financial dimension to prevail over the other dimensions.

In this perspective, we can say that the one-dimensionality of the ‘hierarchization’ of performance can lead organizational malfunctions, which is damaging to an unwritten but crucial principle tied to the systemic vision highlighted above: the notion of ‘loyal collaboration’ between institutions and public bodies, which is fundamental in order to achieve common ends and objectives of a system. The notion of loyal collaboration has always found limited application in the entire Italian public sector to the point of making it complex, if not impossible, to develop clear and consolidated relationships between institutions. It should be reconsidered and fully included among the essential principles for the positive functioning of the public administration, at the same level of efficiency, effectiveness, and economy. To remain in the administrative perspective and to better understand the weight of the relationships, we could ask ourselves: What would happen to any group if the companies that are part of it were not motivated by the unwritten notion of loyal collaboration? They would simply fail or they would be destined to a rapid decline, and their very reason for aggregating in the first place would no longer exist. This is precisely the risk that public entities and public institutions expose themselves to when they are reluctant or incapable of respecting the aforementioned principle, which is fundamental to ensure public value and for the construction of common good, within individual public organizations and within the whole public administration of reference.

In light of this, we can therefore affirm that the lack of loyal collaboration has amplified the negative effects produced by the phenomenon of ‘hierarchization’ of performance and has ultimately resulted in the creation of a vicious circle.

The legal reforms that came about in following years, albeit with different motivations and very different aims, never guaranteed a true, harmonious, and balanced approach to achieve their stated objectives and favored emphasis on financial measurement above all others.

The Brunetta reform, as mentioned above, is particularly useful from a cultural and formative perspective because it incorporates a series of values that, for the first time, are not integrated in a merely abstract list of the single elements. However, it is important to note that the norm does present some particularly significant critical points. The crucial relationship between policy and management (Bianchi and Rivenbark 2014) is not well addressed, especially in terms of the definition and planning of objectives that are compatible with the previous phases of strategic planning.

In any case, the Brunetta reform should have implicitly led to a requalification or an improvement of the relationship between policy and management. It is precisely this relationship between policy and management that should have benefited from the reform by capitalizing on what are considered its key elements (Ricci and Serluca 2013):

- reflection on the identity and mission of a given institution;
- formal procedures for decision making;
- the definition of strategic objectives;
- the use of performance management tree.

Furthermore, it does not address the topics of co-value, co-production, and co-creation of public value which, as mentioned above, is fundamental to public administration (Benington 2011). On the other hand, focusing on the concept of public value as the ultimate aim of the actions of public organizations, rather than on economic value, limits the process of ‘hierarchization’ of performance.

What follows is that measurements based purely on financial performance cannot capture the many innovations related to production, organization and even culture that have taken place in recent years in the public sector. Within the complex relationship between institutions and citizens, the potential qualitative contribution of active engagement by civic stakeholders and by the direct recipients of the services offered by the public administration is more important than ever, although it is not always easy to highlight in terms of public value generated.

On this topic:

‘Performance measurement enables officials to hold organizations accountable and to introduce consequences for performance. It helps citizens and customers judge the value that government creates for them. And it provides managers with the data they need to improve performance’ (Osborne and Plastrik 2000, p. 247).

For example, consider the responsible use of water or the act of recycling diligently. Paradoxically, this also holds true in the measurement of stakeholder engagement (typically non-financial) as well, which the public sector needs more than ever to understand the implementation of its action plans and to verify the leadership capabilities of its management team and political actors. Naturally, this requires institutions that are more open, dynamic, and able to truly interact with stakeholders and citizens who are willing to actively participate in public life. These represent the key characteristics or requirements of the individuals involved.

In conclusion, we can affirm that performance is of crucial importance if it is considered a tool and not an end result. In this sense, performance can be correctly defined as a measure of financial compatibility of one or more priorities (public value, social, environmental, etc.). Conversely, when performance is seen as the aim of a public administration, there is the risk that it may not guarantee balanced outcomes and could even result in the destruction of public value (Esposito and Ricci 2015).

It is indisputable that financial performance is a clear indicator of managerial balance, considering the activities typically carried out by the public administration and its ultimate ends. However, it is also obvious that this and all other evaluations based on the same ‘criteria’ are very limiting because they presuppose a set up that does not give enough consideration to the multidimensionality previously discussed. Furthermore, and particularly for our purposes, this type of financial one-dimensionality is in conflict with the remaining dimensions which are

explicative of performance. This conceptual understanding can only be rejected: As was mentioned previously, it is necessary to adopt a systemic logic of performance in the public administration. Every form of ‘compartmentalization’ leads to evident ideological conflicts which appear to be paradoxical and in conflict with a complete and global evaluation of management of public organizations. The examples shown above, in this sense, allow us to underline the need for deep and complete reflections on the topic of performance in the public administration, which cannot be structured based on the compartmentalized, exclusionary, and oppressive logics outlined above. Conversely, systemic approaches which consider the coexistence of many dimensions can reconcile the obvious and physiological specificities within every dimension, allowing for the optimization of the various levels of performance and the achievement of an overall balance within the public administration (Guthrie and English 1997; Bouckaert and Halligan 2008).

The continuous legal reforms of recent years, especially in Italy, have progressively accentuated the ‘hierarchization’ of performance through interventions aiming at rationing (rather than rationalizing) the financial resources of public organizations. Considering the extremely limited results achieved in terms of increasing efficiency and effectiveness, the improvement of the financial dimension of performance—still necessary in the systemic view outlined above—could be pursued through a different approach: An improvement in the ability to forecast could, in fact, represent one of the possible solutions to guarantee a true optimization without the disadvantages that are implicit to the one-dimensional solutions illustrated thus far.

As a final reflection on public sector developments in Italy, we can affirm that the country has been affected by an excessive reliance on legislative reforms. Performance management reforms have been imposed without periodically evaluating the results achieved, and rules have been changed often without leading to real institutional change. It is necessary to identify the causes of problems within a given community, country, and public administration, rather than to merely measure their financial or quantitative impact. In this sense, Behn (2003, p. 595):

The real, ultimate outcome that citizens seek from our public schools is children who grow up to become productive employees and responsible citizens. But using a measure of employee productivity and citizen responsibility to motivate performance creates a number of problems. First, it is very difficult to develop a widely acceptable measure of employee productivity (do we simply use wage levels?), let alone citizen responsibility (do we use voting participation?). Second, schools and teachers are not the only contributors to a future adult’s productivity and responsibility. And third, the lag between when the schools and teachers do their work and when these outcomes can be measured is not just months or years, but decades.

Western economies are now well aware of the political and managerial problems facing their public institutions, although with some delay, they have found ways to resolve these problems by taking into consideration multiple factors and not only financial ones (Ricci and Serluca 2013).

References

- Adams, C. A., Muir, S., & Hoque, Z. (2014). Measurement of sustainability performance in the public sector. *Sustainability Accounting, Management and Policy Journal*, 5(1), 46–67.
- Aguinis, H., Joo, H., & Gottfredson, R. K. (2011). Why we hate performance management—and why we should love it. *Business Horizons*, 54(6), 503–507.
- ANAC (2013). *Relazione annuale sull'attività svolta—Anno 2013*. Roma.
- Anderson, G. F., & Frogner, B. K. (2008). Health spending in OECD countries: Obtaining value per dollar. *Millwood: Health Aff*, 27, 1718–1727.
- Atkinson, A. A., Waterhouse, J. H., & Wells, R. B. (1997). A stakeholder approach to strategic performance measurement. *Sloan Management Review*, 38(3), 25–37.
- Barzelay, M. (1992). *Breaking through bureaucracy: A new vision for managing in government*. Berkeley: University of California Press.
- Beck, J. T., & Bozeman, B. (2007). Public values: An inventory. *Administration & Society*, 39(3), 354–381.
- Behn, R. D. (2001). *Rethinking democratic accountability*. Washington, DC: Brookings Institution Press.
- Behn, R. D. (2003). Why measure performance? Different purposes require different measures. *Public Administration Review*, 63(5), 586–606.
- Benington, J. (2011). From private choice to public value? In J. Benington & M. Moore (Eds.), *Public value: Theory and practice*. Basingstoke: Macmillan.
- Berman, E. (2002). How useful is performance measurement. *Public Performance & Management Review*, 25(4), 348–351.
- Bianchi, C., & Rivenbark, W. C. (2014). Performance management in local government: The application of system dynamics to promote data use. *International Journal of Public Administration*, 37(13), 945–954.
- Bigoni, M., & Deidda, G. E. (2013). Brunetta's reform swan song? An assessment of its success in local governments through the analysis of its tools. *European Scientific Journal*, 9(29), 403–428.
- Borgonovi, E. (2001). Il concetto di valore pubblico. *Azienda Pubblica*, 14(2/3), 185–188.
- Borgonovi, E. (2009). Il diritto dovere di valutare e premiare il merito nelle amministrazioni pubbliche. *Azienda Pubblica*, 22(2), 199–204.
- Borgonovi, E., & Compagni, A. (2013). Sustaining universal health coverage: The interaction of social, political, and economic sustainability. *Value in health*, 16, S34–S38.
- Borgonovi, E., & Valotti, G. (2009). Riforme istituzionali e management: Un'alleanza per il cambiamento. *Azienda Pubblica*, 3, 403–407.
- Bouckaert, G., & Halligan, J. (2008). *Managing performance: International Comparison*. London: Routledge.
- Bovaird, T. (1996). The political economy of performance measurement. In A. Halachmi & G. Bouckaert (Eds.), *Organizational performance and measurement in the public sector: Toward service, efforts and accomplishment reporting*. Westport, CT: Quorum Books.
- Bovens, M., Goodin, R. E., & Schillemans, T. (Eds.). (2014). *The Oxford handbook of public accountability*. Oxford: Oxford University Press.
- Bozeman, B. (2007). *Public values and public interest: Counterbalancing economic individualism*. Washington, DC: Georgetown University Press.
- Bracci, E., Deidda, G., Gagliardo E., & Bigoni, M. (2014). Performance management systems and public value strategy: A case study. In J. Guthrie, G. Marcon, S. Russo, & F. Farneti (Eds.), *Public value management, measurement and reporting* (pp. 129–157). Bingley: Emerald Group Publishing Limited.
- Cambridge Dictionary. (2016). 'Hierarchy' entry. Cambridge: Cambridge University Press.
- Campesi G. (2011). *The Arab Spring and the Crisis of the European Border Regime: Manufacturing Emergency in the Lampedusa Crisis*. EUI Working Paper, RSCAS 2011/59 Mediterranean Programme.

- CiVIT (2011). *Analisi dei sistemi di misurazione e valutazione della performance*. CiVIT.
- CiVIT (2012). *Analisi dei sistemi di misurazione e valutazione della performance*. CiVIT.
- Cuganesan, S., Jacobs, K., & Lacey, D. (2014). Beyond new public management: Does performance measurement drive public value in networks? In J. Guthrie, G. Marcon, S. Russo, & F. Farneti (Eds.), *Public value management, measurement and reporting* (pp. 21–42). Bingley: Emerald Group Publishing Limited.
- De Bruijn, H. (2007). *Managing performance in the public sector*. New York: Routledge.
- De Lancer, J. P., & Holzer, M. (2001). Promoting the utilization of performance measures in public organizations: An empirical study of factors affecting adoption and implementation. *Public Administration Review*, 61, 693–708.
- Del Bene, L. (2014). L'applicazione del D.Lgs 150/2009 negli enti locali tra opportunità e rischi. L'esperienza del Comune di Livorno. *Azienda Pubblica*, 1, 39–56.
- Drucker, P. F. (1954). *The practice of management*. New York: Harper Business.
- Drucker, P. F. (1976). What results should you expect? A user's guide to MBO'. *Public Administration Review*, 36, 12–19.
- Dubnick, M. (2005). Accountability and the promise of performance: In search of mechanisms. *Public Performance and Management Review*, 28, 376–417.
- Dunleavy, P., & Hood, D. (1994). From old public-administration to new public management. *Public Money & Management*, 14(3), 9–16.
- Epstein, M. J., & Birchard, B. (2000). *Counting what counts: Turning corporate accountability to competitive advantage*. Cambridge, MA: Perseus Books.
- Esposito, P., & Ricci, P. (2015). How to turn public (dis)value into new public value? Evidence from Italy. *Public Money & Management*, 35(3), 227–231.
- Fryer, K., Anthony, J., & Ogdan, S. (2009). Performance management in the public sector. *International Journal of Public Sector Management*, 22(6), 478–498.
- Galli D. (2012). *Indicatori e ambiti di misurazione*, in VV.AA. *Quale performance? Indicatori e sistemi di misurazione in uso nei Ministeri di sette Paesi OCSE, Proposte per le amministrazioni centrali italiane*, OCAP White Paper, n. 2/2012, Milano: EGEA.
- Galli D. & Turrini A (2013). *Il performance management e le amministrazioni pubbliche italiane, White Paper "La PA che vogliamo. Proposte per un cambiamento possibile"*, 85–103.
- Guthrie, J., & English, L. (1997). Performance information and programme evaluation in the Australian public sector. *International Journal of Public Sector Management*, 10(3), 154–164.
- Guthrie, J., Humphrey, C., Jones, L., & Olson, O. (Eds.). (2005). *International public financial management reform: Progress, contradictions and challenges*. Greenwich, Connecticut: Information Age Publishing.
- Halachmi, A. (2005). Performance measurement is only one way of managing performance. *International Journal of Productivity and Performance Management*, 54(7), 502–516.
- Hatry, H. P. (2002). Performance measurement: Fashion and fallacies. *Public Performance & Management Review*, 25(4), 352–358.
- Holmblad, Brunsson K. (2002). Management or Politics-or Both? How management by objectives may be managed: A Swedish example. *Financial Accountability & Management*, 18(2), 189–209.
- Holzer, M., & Yang, K. (2004). Performance measurement and improvement: An assessment of the state of the art. *International Review of Administrative Studies*, 70(1), 15–31.
- Hood, C. (1991). A public management for all seasons? *Public Administration*, 69(1), 3–19.
- Hughes, O. E. (2003). *Public management and administration: An introduction*. Basingstoke: Palgrave Macmillan.
- Jones, R., & Pendlebury, M. (2000). *Public sector accounting* (5th ed.). London: Pearson Prentice Hall.
- Jones, R., & Pendlebury, M. (2010). *Public sector accounting* (6th ed.). London: Pearson Prentice Hall.
- Kaplan, R. S., & Norton, D. P. (2001). Strategic performance measurement and management in nonprofit organizations. *Nonprofit Management & Leadership*, 11(3), 87–104.

- Kelly G., Mulgan G., & Meurs S. (2002). *Creating public value. An analytical framework for public service reform*, UK: Cabinet Office. <http://www.strategy.gov.uk>.
- Kloot, L., & Martin, J. (2000). Strategic performance management: A balanced approach to performance management issues in local government. *Management Accounting Research*, 11(2), 231–251.
- Lapsley, I., & Mitchell, F. (1996). *Accounting and performance management*. London: Paul Chapman Publishing.
- Lemieux-Charles, L., McGuire, W., Champagne, F., Barnsley, J., Cole, D., & Sicotee, C. (2003). The use of multilevel performance indicators in managing performance in health care organizations. *Management Decision*, 41(8), 760–770.
- Liguori, M., Sicilia, M., & Steccolini, I. (2012). Some like it non-financial...Politicians' and managers' views on the importance of performance information. *Public Management Review*, 14(7), 903–922.
- Marchetti, C. (2014). Rifugiati e migranti forzati in Italia. Il pendolo tra 'emergenza' e 'sistema'. *REMHU-Revista Interdisciplinar da Mobilidade Humana*, 43, 53–70.
- Modell, S. (2004). Performance measurement myths in the public sector: A research note. *Financial Accountability & Management*, 20(1), 39–55.
- Monteduro, F. (2010). Il ciclo di gestione delle performance. In M. Barbieri, L. Hinna, S. Mameli, F. Monteduro, & G. Valotti (Eds.), *Gestire e valutare le performance nella PA*. Maggioli: Rimini.
- Monteduro F. & Hinna L. (2007). *La misurazione delle performance: Cos'è e come utilizzarla nelle decisioni*, in VV.AA., *Misurare per decidere: La misurazione della performance per migliorare le politiche pubbliche ed i servizi*, Dipartimento della Funzione Pubblica, Soveria Mannelli: Rubbettino.
- Moore, M. H. (1995). *Creating public value. Strategic management in government*. Cambridge, MA: Harvard University Press.
- O'Flynn, J. (2007). From new public management to public value. *Australian Journal of Public Administration*, 66(3), 353–366.
- OECD. (1997). *In search of results: Performance management practices in 10 OECD countries*. Paris: Oecd Publishing.
- OECD. (2004). *Public sector modernization: Governing for performance*. Paris: Policy Brief, Oecd Publishing.
- OECD. (2005). *Modernising government. The way forward*. Paris: Oecd Publishing.
- Osborne, D., & Plastrik, P. (2000). *The reinventor's fieldbook: Tools for transforming your government*. San Francisco, CA: Jossey-Bass.
- Otley, D. (1999). Performance management: A framework for management control systems research. *Management Accounting Research*, 10, 363–382.
- Peck, J. (2014). Pushing austerity: State failure, municipal bankruptcy and the crises of fiscal federalism in the USA. *Cambridge Journal of Regions, Economy and Society*, 7, 17–44.
- Pollitt, C., & Bouckaert, G. (2004). *Public management reform: A comparative analysis*. Oxford: Oxford University Press.
- Riccaboni, A. (1993). Gerarchia degli obiettivi d'impresa e sistema di gestione delle performance. In AAVV (Ed.), *Scritti in onore di Carlo Masini*. Egea: Milano.
- Ricci, P. (2016). Accountability. In A. Farazmand (Ed.), *Global encyclopedia of public administration, public policy and governance*. Switzerland: Springer International Publishing.
- Ricci, P., & Serluca, M. C. (2013). Gli enti locali e la valutazione della performance: Uno stato dell'arte. *RU—Risorse Umane nella Pubblica Amministrazione*, 1, 13–25.
- Ridley, C. E., & Simon, H. A. (1943). *Measuring municipal activities: A survey of suggested criteria for appraising administration*. Chicago: The International City Managers' Association.
- Rivenbark, W. C. (2007). A manager's toolbox. In C. Stenberg & S. L. Austin (Eds.), *Managing local government services*. Washington, DC: ICMA.
- Romano, M., & Cirillo, A. (2015). La misurazione delle performance economico-finanziarie delle università pubbliche italiane. Logiche e strumenti del nuovo sistema di contabilità. *Azienda Pubblica*, 4, 377–394.

- Stigler, G. T. (1971). The theory of economic regulation. *The Bell Journal of Economic and Management Science*, 2(1), 3–21.
- Stoker, G. (2006). Public value management: A new narrative for networked governance? *American Review of Public Administration*, 36(1), 41–57.
- Streib, G. D., & Poister, T. H. (1999). Assessing the validity, legitimacy, and functionality of performance measurement systems in municipal governments. *American Review of Public Administration*, 29(2), 107–123.
- Thiel, S., & Van Leeuw, F. L. (2002). The performance paradox in the public sector. *Public Performance & Management Review*, 25(3), 267–281.
- Thomas G. (2007). *Why is Performance-Based Accountability so Popular in Theory and Difficult in Practice?* Paper presented at World Summit on Public Governance: Improving the Performance of the Public Sector—Taipei City.
- Trencher, G., Yarime, M., McCormick, K. B., Doll, C. N. H., & Kraines, S. B. (2014). Beyond the third mission: Exploring the emerging university function of co-creation for sustainability. *Science and Public Policy*, 41, 151–179.
- Van Dooren W. (2006). *Performance measurement in the Flemish Public Sector: A supply and demand approach*, *Faculteit Sociale Wetenschappen—Onderzoekseenheid: Instituut voor de Overheid*. België: K.U. Leuven.
- Van Dooren, W., Bouckaert, G., & Halligan, J. (2010). *Performance management in the public sector*. New York: Routledge.
- Wang, J., Huapu, L., & Peng, H. (2008). System dynamics model of urban transportation system and its application. *Journal of Transportation Systems Engineering and Information Technology*, 8(3), 83–89.

Part III
Enhancing Governance and
Inter-institutional Coordination through
Outcome-Based Performance Management

Chapter 8

Measuring Coordination and Coherence: Assessing Performance Across the Public Sector

Carmine Bianchi and B.G. Peters

8.1 Introduction

This chapter will bring together two bodies of the literature addressing some of the most important issues in contemporary public administration and governance: policy coordination and performance management. Beginning with the Blair government's interest in "joined-up government" (Bogdanor 2005) and continuing with initiatives such as the "Whole of Government in Australia", contemporary governments have been attempting to create more integrated and coherent approaches to policy problems (Bouckaert et al. 2010). This concern with improving coordination is, at least in part, a reaction to the fragmentation of the public sector during the period of dominance of the New Public Management, but some issues of policy coordination have been present since the beginnings of modern government (see Peters 2015).

The second body of the literature concerns performance measurement and performance management in the public sector (Bianchi and Rivenbark 2013). The use of these performance measurements also has been one of the standard recommendations of the New Public Management, arguing that individuals and organizations within government should be made more accountable for their actions. This accountability is to be achieved through having clear and quantifiable targets for

C. Bianchi (✉)

Department of Political Sciences and International Relations, CED4 System Dynamics Group, University of Palermo, Palermo, Italy
e-mail: bianchi.carmine@gmail.com

B.G. Peters

Department of Political Science, University of Pittsburgh, Pittsburgh, USA
e-mail: bgpeters@pitt.edu

© Springer International Publishing AG 2018

E. Boronovi et al. (eds.), *Outcome-Based Performance Management in the Public Sector*, System Dynamics for Performance Management 2, DOI 10.1007/978-3-319-57018-1_8

143

their actions, measuring the extent to which those targets are attained, and then punishing or rewarding the actors involved accordingly (Bouckaert and Halligan 2008). The use of these forms of measurement replaces conventional forms of accountability with more quantifiable and possible less-politicized forms.

Although performance measurement has been at least partially successful in assessing how well individual organizations and their members are performing their tasks in government, that focus on the individual organization has tended to some extent to exacerbate the problems of coordination. If managers must be concerned primarily with meeting a set of targets for their own organization, they are less likely to cooperate with other organizations and to take a more extensive vision of the performance of the public sector.

We will develop these two strands of concerns in public management separately, and then bring them together to demonstrate how improving performance management can be used to enhance levels of coordination, and vice versa. And we will further demonstrate the utility of a more dynamic approach to performance management than is conventionally used in both assessing performance and in contributing to coordination.

8.2 Measuring Coordination in the Public Sector

To begin to develop performance measures for coordination, we must first think about ways of measuring coordination itself. Unfortunately, despite the centrality of this concept in discussions of public administration and public policy for many decades, the measurement of coordination has been relatively weak. Coordination appears to be very much like Justice Potter Stewart's conception of pornography—he could not define it but he knew it when he saw it.

There are, however, some qualitative indicators of coordination. The most notable of these is Metcalfe's (1994, see Table 8.1) nine-point scale ranging from independent decision making by ministers to a clear government strategy. While each of the points along the scale is sensible and represents a real variant of level of coordination success, the discussion provides no clear indicators of how to classify particular situations. The application of this scale to real-world coordination events, therefore, depends upon the judgment of the individual applying it.¹

Braun (2008) developed a somewhat less complex scale of coordination (see Table 8.2), having only five points, going from no coordination through to strategic coordination. While less complex than Metcalfe's approach, this model of coordination assumes that these levels of coordination constitute a Guttman scale, with achieving one step assumed to be essential to achieving the subsequent step.

¹There are, of course, means of attempting to strengthen the use of judgment such as using multiple coders and testing inter-coder reliability, but the method still ultimately relies on judgment, whether expert or not.

Table 8.1 Metcalfe's policy coordination scale

9.	Government strategy
8.	Establishing central priorities
7.	Setting limits on ministerial action
6.	Arbitration of policy differences
5.	Search for agreement among ministers
4.	Avoiding divergences among ministers
3.	Consultation with other ministers (feedback)
2.	Communication with other ministers (information exchange)
1.	Independent decision making by ministers

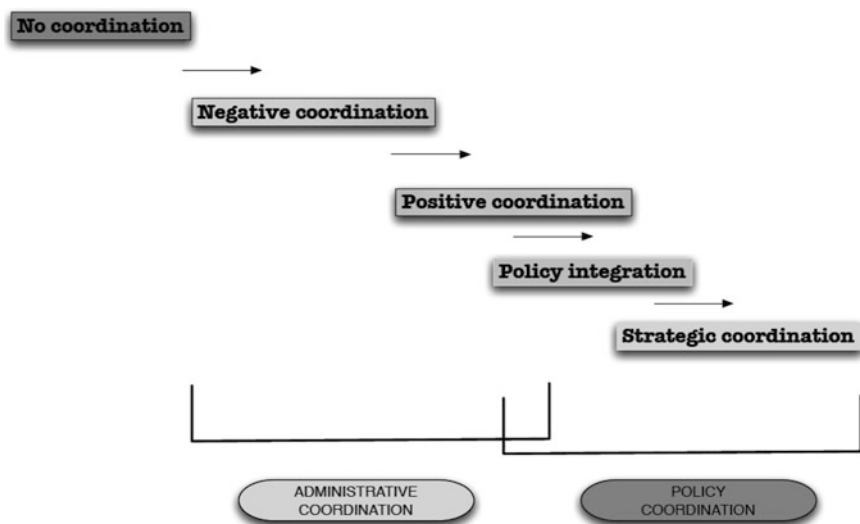
We can consider the distinctions among positive, negative, and strategic levels of coordination based on, and extending, Scharpf's (1997) analysis of policy coordination.² These conceptions of policy coordination have been rather abstract and theoretical, and therefore we must consider how to convert those approaches into more operational indicators of performance in creating more concerted public action. These categories also will depend upon applying judgment on the part of the researcher who is attempting to understand existing levels of coordination.

Another means of attempting to measure coordination in the public sector is through examining the mechanisms within the public sector that are dedicated to producing more coordination and integration. For example, in our study of coordination and specialization in seven industrial democracies, we (Bouckaert et al. 2010) mapped the creation and dissolution of both structures and processes concerned with coordination. Other efforts of the same sort have documented the ebb and flow of government efforts to generate coordination (see Schout and Jordan 2005) in the European Union and in individual countries.

The movement toward "joined-up government", beginning with the Blair government in the UK and then diffusing to other countries, contained one set of instruments designed to generate more coherence in the public sector (Bogdanor 2005; Pollitt 2003). These efforts at greater coordination in governance—institutional as well as substantive—tended to focus on the development of hierarchical controls created by the center of government. And, whether or not called "joined up", this general change in governing has been found in a number of countries (Dahlström et al. 2010).

Policy integration is sometimes discussed as equivalent to coordination, although it implies attempts to make policies coherent *ex ante*, rather than getting programs to work together *ex post* (see Jordan and Lenschow 2010). In addition to the above measures of coordination, Briassoulis (2005) developed an implicit scale for efforts (Table 8.3) intended to create greater policy integration in environmental

²By negative coordination Sharpf meant that organizations were aware of each other's activities and goals, and attempting not to conflict. Positive coordination, on the other hand, implies that rather than simply avoiding conflicts the organizations would attempt to work together. And finally strategic coordination would mean working together toward broader, systemic goals.

Table 8.2 Braun's Guttman scale of coordination

policy. Unlike the catalogs of structures and processes above, this is a more comprehensive listing of instruments that can be used to produce more coherent policies. It includes some structural elements, but also has some elements that resemble Howlett's (2000) concept of procedural instruments for governing. In other words, coordination is created through processes (including performance management) as well as through formal structures and institutions.

As well as attempts to coordinate across organizations within the public sector, there may also be needs to coordinate across governments, and to create more integrated patterns of service delivery, especially in metropolitan areas with multiple smaller governments. This coordination is sometimes handled through the creation of special districts, e.g., the Port Authority of New York, that provide services such as transportation that have an extensive geographical scope. The problem may also be addressed through creating metropolitan area governments. For example, in Italy, law has defined metropolitan cities since the year 2014. Such institutions will replace the Italian provinces, which have been the third administrative level, below the State and Regions. Each metropolitan city is led by a metropolitan mayor, supported by a Metropolitan Council and by a non-legislative assembly (the metropolitan conference). Members of the Metropolitan Council are elected and chosen by mayors and city councilors of each municipality in the metropolitan city. The metropolitan mayor is the mayor of the former capital of the province.

In addition to the qualitative and structural indicators of coordination, there have been more empirical attempts at measuring coordination. While most qualitative indicators provide either a descriptive set of criteria or examine the potential for

Table 8.3 Briassoulis's dimensions of policy integration

Institutional criteria
Legislative criteria
Administrative criteria
Financial measures
Market measures
Technical measures
Communication measures
Hybrid measures

coordination, these more quantitative indicators do attempt to assess the extent to which coordination is actually being achieved. Even here, however, attempting to assess what is actually happening with policies may be difficult, and these measures tend to measure interactions among organizations and individuals more than they actually measure the integration of policies.

Perhaps the measures of policy coordination that has come closest to actually assessing levels of coordination are those developed by Jennings and Crane (1994) in their study of employment programs in the American states. Although they were not able to measure coordination per se, their interviews with program administrators did reveal the perceptions of the level of coordination that existed, as well as the causes for success and failure. Interestingly, given all the concern with program design, the major factors associated with the success of coordination efforts were interpersonal relationships among the administrators and leadership within the organizations.

In a later article, Jennings and Ewalt (1998) presented a list of some 41 possible indicators of coordination activities by state level. Most of these were measures of activity and attempts to generate coordination, rather than measures of any success in those efforts. These attempts to coordinate were related to a number of organizational characteristics, as well as to the characteristics and behaviors of the leadership of the organizations.

Another attempt at measuring coordination is premised on the central role of information in the public sector, and the tendency of public organizations to hoard information (see Husted and Michailova 2002).³ Willem and Buelens (2007; see also Boateng and Agyemang 2016) used questionnaires to measure the extent to which organizations in the Belgian government shared their information. Although they did find some structural effects on coordination, the major factors associated with knowledge sharing were interpersonal and cultural.

New Zealand provides an interesting context where intensive reforms have been implemented since the beginning of the 1990s, in order to foster coordination between policy design and implementation. This can be considered as a pivotal case

³One classic case of information hoarding occurred in the American federal bureaucracy when various intelligence organizations did not share information prior to the 9/11 disaster. See Peters (2015).

of vertical coordination, implying the search of a connection between strategic goals (strategic result areas, set by elected officials) and related departmental objectives (key result areas, set by administrators). Also, horizontal coordination is a main issue of such reforms. To foster coordination, nine broad policy areas were identified by the New Zealand reformers: economic growth, enterprise and innovation, external linkages, education and training, community security, social assistance, health and disability services, treaty claims settlement, and environment (Boston and Pallott 1997).

The need of a strategic thinking and a strategic conversation in policy design and implementation was emphasized as a major reason for reforming the public sector. To enhance “strategic conversation” throughout and across public administration, the term *purple zone* was coined (Matheson et al. 1997; Alford et al. 2016; Elliot 1998; Shergold 1997). This is a blurred area of decision making and responsibility, underlying a degree of indeterminacy on the roles and relationships between the political and administrative domain. According to this approach, the relations between elected officials and public servants should be described over a continuum, rather than being conceived as sharply separated (Svara 2001). The dynamic complexity characterizing decision making is a major reason requiring such perspective. The “purple zone” is “an amalgam of separation and integration, in which the benefits arising from distinctive responsibilities are complemented by those flowing from strategic coherence and common “branding”” (Matheson et al. 1997, p. 5). An incremental and design (i.e., learning-oriented), rather than structured, approach in decision making is claimed to cope with dynamic complexity and problem wickedness.

In this chapter, we will discuss the possibility of developing performance management systems that can assess coordination and coherence among public organizations and their policies, as well as the performance of individual organizations. We will begin with Metcalfe’s (1994) scale of coordination, and we will also consider attempts by governments such as that of New Zealand to develop coordination measures and strategic management measures, as one component of their performance management system and its strategic results areas.

8.3 Examples of Coordinated Performance

We will discuss the development of these indicators of coordinated performance in general terms, but also consider them in relationship to two policy areas. These two policy areas represent some interesting contrasts in the challenges they present to policy coordination and integration, and in turn they, therefore, present interesting challenges for measuring the success of coordination.

8.3.1 *Social Policy and Health*

The first broad policy domain we will consider is health and social policy. To be effective, these two policies must have coherent and compatible policies, and must make attempts to serve the “whole client”. It may be difficult to provide adequate health care for individuals without understanding their social situation, and vice versa. While the need to coordinate these policies has been recognized for decades, effective coordination continues to elude most governments. That said, we can still consider the mechanisms for pursuing that goal and develop the means of assessing performance on coordination.

One of the earlier efforts to examine the manner in which these policies worked together, or did not, was in the Model Cities program in the USA in the 1960s. One of the fundamental ideas behind this attempt to revitalize blighted areas in American cities was to integrate the services being provided to residents of the “model neighborhoods”. While there were a large number of public and private organizations providing services to these areas, there was little or no attempt to coordinate their activities and to present more or less coherent packages of services to citizens.

A survey of agencies operating in these neighborhoods identified almost sixty service providers, most working with social and health policies of some sort, but there was almost no discussion among the organizations, even those that nominally were providing similar types of services. Some clients would be contacting multiple organizations attempting to receive the same types of services, but those organizations were not working together to provide the services. And there was even less cooperation across policy sectors (Main et al. 1972). Efforts to improve levels of referral across agencies led to some increase of referrals, but mainly *within* policy sectors rather than across.

At approximately the same time, the UK embarked upon a major effort to produce greater integration of social and health policies. As documented by Challis et al. (1988), this effort was an attempt to produce coordination across the entire social policy sector. As has been true for many such efforts, the usual organizational and political barriers to effective coordination within the public sector won out over the hopes of the policy planners and coordinators.

The two examples above may appear like prehistory, but much of the same problem of coordination within this policy exists today. For example, even in well-organized welfare states when patients leave hospital, especially elderly patients living alone, there is often no connection with the social services they will need for care during their convalescence, and the patients may have to return to the hospital (Leichsenring 2004). These failures of coordination are perhaps most acute for children, with frequent failures of health organizations failing to report suspected abuse to social workers and the police (Marinetti 2011).

It is easy to catalog a host of failures in the coordination of social and health policy. It is substantially more difficult to develop measures of the levels of coordination that may or not exist. And if we can provide those measures, then can

we develop measures for the performance of individuals and organizations in generating that coordination. Assessing coordination will go a long way toward the additional goal of being able to attribute that coordination to the behaviors of individuals and/or their organizations.

First, we can measure the degree of contact among actors. For example, are their coordinated care conferences around individual patients or more generally for planning a more seamless transition of clients in and out of health care and social care? The difficulties of arranging this seamless movement of clients may be exacerbated when organizations are operating at different levels of government, or when those organizations have different catchment areas (Christensen et al. 2014).

Similarly, is there any significant level of knowledge sharing among organizations, whether about individuals or about more general policies? As noted, information hoarding is a common form of avoiding coordination, and to some extent client hoarding can accomplish some of the same purposes—especially in social and health policy. When organizations have clients, they have some claims to power and can attract resources (see Cook and Cheshire 2015), and they can trade potential access to clients for influence with other organizations.⁴

We can also consider structural solutions for coping with the problems of coordination and policy integration between health and social services. One common example is creating “superdepartments” in which a range of linked programs may be included under a single ministerial roof. While there has been some limited success with structures of this type, in most instances the coordination issues are merely made intramural rather than extramural, but they persist and may become even stronger because the various policy sectors are competing for primacy within the one department.

In Italy, coordination between social and health policies should be pursued by metropolitan cities through “social policies and health” services. Such services should support coordination between municipalities and the regional health care system (particularly with the health care agencies). At both regional and State levels, such coordination should be mainly pursued at corporate level, i.e., through the strategic planning process led by the head of cabinet. Though overlapping areas exist between health care and welfare agencies, separate competences are assigned in both the Italian regions and the State at ministry or agency level. Policy coordination in terms of social and health services between the State and regional level can be managed through conference of services; however, such coordination instruments have been characterized by a static and short-term—rather than outcome-oriented—perspective.

We could continue adding examples of activity measures to this few already mentioned. But these activity measures do not address the more significant question of policy integration. Individuals could participate in coordination activities but those could fail, given the numerous barriers that exist to coordination. Therefore, we need to develop measures that assess the extent to which individual clients are

⁴On the resource dependency model considered more generally see Davis and Cobb (2010).

indeed provided with more or less connected services. For example, we might look at hospital readmissions for discharged patients, especially elderly patients, to assess the extent to which the social network that is meant to serve them after discharge is functioning.⁵

8.3.2 *Food Policy*

The other policy area we will consider is coordination in food policy. In order to be able to deliver pure food from the farm to the diner's plate requires coordination across organizations and across time. These organizations may be focused on several different policy domains—agriculture, health, and consumer safety—and functioning at different levels of government. This dynamic chain coordination is somewhat different from the usual static conception of interorganizational coordination. Successful food policies depend on adequate performance at each stage of the process, but also require the linking of these separate organizations as the food goes from farm to consumer.

While the above coordination issues in the regulation of food quality are all to some extent associated with agriculture and food per se, food and agriculture issues must be associated with policy areas such as energy and the environment. For example, intensive agriculture that may be useful for producing the large quantities of food is energy intensive and also can have severe environmental effects (Grochowska 2014). And with biomass as a potential source of energy, land may be diverted from food production to making ethanol or other renewable forms of energy. Producing energy may be more profitable for the landowners who will have to worry less about spoilage, and also they can use the entire plant for production.⁶

Whereas the coordination of health and the social services is perhaps primarily a problem of implementation, and may need to be resolved at the lower levels of organizations, the problems of coordination in food policy may be more those of policy integration. That is, to consider adequately the question of providing sufficient and high-quality food supplies at affordable prices requires the involvement of energy, environment, land use, and increasingly, international policy actors.⁷

Given the need to focus more on policy than on service delivery with food policy, the measures of coordination will need to focus more on the nature of the policies being implemented rather than implementation per se. Therefore, the measures of success can be less on activity than on the extent to which policies do

⁵One small component of the Affordable Care Act in the United States penalizes hospitals for high rates of readmission. But the problem may be less with the hospital itself than with the linkages to social services.

⁶That is, for corn in particular not only are the ears of corn used to make ethanol but the stalk and all.

⁷Very few countries are now self-sufficient in food production so trade and aid are essential elements of food policy.

in fact take into account needs and priorities in other policies sectors. That more integrated legal framework will still, of course, have to be implemented but producing a food policy that does indeed consider a range of alternative policy concerns will be the necessary first step to success.

The above discussion has been premised on a notion of food policy as primarily directed toward the consumer of food. For farmers and food processors, on the other hand, the important goals of the policy area may be rather different. Those different perspectives held by producers and consumers may be common in many policy areas, they also will constitute a locus for considering greater policy integration. And even within the producer group, there can be substantially different goals for small holders and for agribusiness firms (Locke and Henley 2016).

The above-mentioned problems within food policy make coordination difficult, which in turn makes the measurement of that coordination equally difficult. The organizations that are relevant for food policy—ranging from social services to international trade—may have very different goals and may actively resist coordination. Even in a relatively well-organized government such as Canada developing a “joined-up food policy” has proven difficult (MacRae 2011), and assessing the level of coordination appears even more difficult.

At a first stage of understanding how organizations may take into account the goals of other organizations in food policy, the existence of coordination structures at a high level within government is necessary (see Peters 2016). These structures tend to be more prevalent in less developed countries, such as those in Africa, given the importance of agriculture in the economies and the threats of food insecurity to the population (Pritchard 2016). For more developed countries, measuring the strength of linkages among organizations in the chains attempting to guarantee food safety from farm to table is also an important means of measuring coordination in food policy.

8.4 Using Dynamic Performance Measurement/Management to Foster Coordination and Coherence

The context for coordination and performance described above provides an ideal conceptual foundation for using a dynamic performance management approach (Bianchi 2016) to deal with the need of coordination and coherence in public policy design and implementation. Most problems requiring that public and private organizations develop a strong collaboration are related to social “wicked” issues. The issues imply a high level of dynamic complexity in the design and implementation of policies leading to sustainable outcomes in the relevant system (Bianchi 2015; Head and Alford 2013; Laegreid and Rykkja 2014; Rittel and Webber 1973, p. 160).

A dynamic and outcome-based performance measurement/management approach is needed particularly in such contexts, since delays and nonlinear feedback relationships affecting policy outcomes require that decision makers be supported to frame the structure and behavior of the systems where policies will be implemented. This approach appears in part in the literature on procedural policy instruments (Howlett 2000). This approach may help them to detect the risks of policy resistance (Ghaffarzadegan et al. 2011), i.e., of unintended effects of implemented policies that may appear consistent, if observed on a static and sectoral perspective, but may generate problems in the long run because of lack of coordination, or lack of adaptation.

In this view, a number of challenges can arise when we attempt to design sustainable and “intelligent” performance management systems—both inside and throughout public sector organizations. *First*, the focus of such systems should not only be on the *end-results*, i.e., the flow—or net change—generated by the implemented policies in a given time span into the initial endowment (stock) of strategic resources that an organization cannot purchase on the market (e.g., financials and intangibles generated by the organizational routines). One should also focus on the *performance drivers*, i.e., those critical success factors for achieving the end-results. Performance drivers should be measured, monitored, and affected in the short run, in order to influence the achievement of targeted outcomes. Performance drivers are gauged as ratios between the current strategic resource levels affecting performance and related benchmarks or desired levels (for instance: “skills/desired skills” ratio, affecting service delivery failure rate). One should also outline the policies to adopt in order to affect the strategic resources (i.e., the stocks of tangible and intangible factors to build up and deploy together with others) that will influence performance drivers, and—through them—the end-results, which will feedback on strategic resources (Fig. 8.1).

Second, the relevant boundaries of performance management systems adopted by public sector organizations should not be limited to an institutional context. An outcome-based view of performance requires the use of proper measures that are able to gauge the long-term and wider impact of the implemented policies by a single player on a system that is usually much wider than the physical and juridical boundaries of a single institution. This implies that an external (i.e., interinstitutional) view of performance should be combined with an internal one, in respect to a single organization (e.g., a municipality). From the interplay of the two perspectives enhancing a strategic dialogue among the key players, the sustainable development of a single institution can be made compatible with the development of the social/economic system to which it belongs. This would further ensure the lifelong endurance of both each individual institution in a system and of the system itself.

The design and use of performance measurement/management systems aimed to foster policy integration and joined service delivery should take into consideration the level of dynamic complexity, that is, intrinsic to the environments where policy makers are expected to interact.

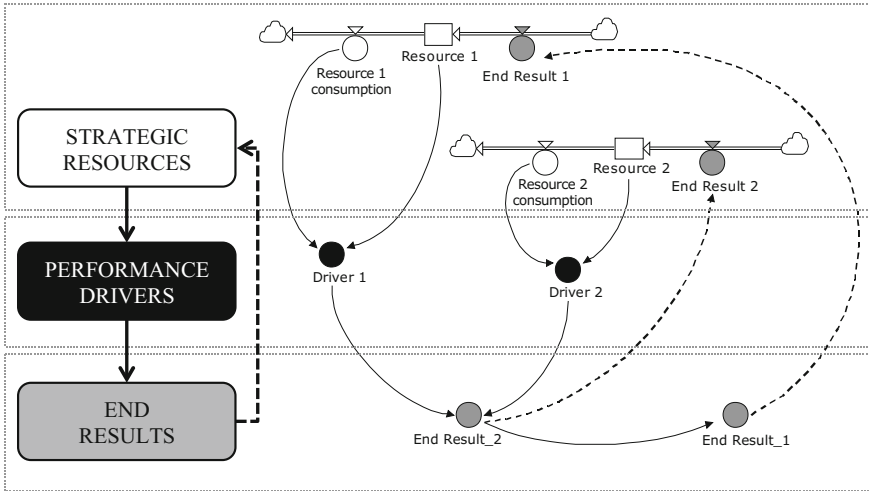


Fig. 8.1 A dynamic performance management view

This dynamic complexity could be measured in relation to a variety of factors, such as:

1. Number of policy/decision makers affecting the relevant system's outcomes;
2. Number of variables to manage and to take into account in policymaking;
3. Number of decision levels that are sequentially connected;
4. Delays affecting the system's outcomes;
5. Diversity of issues characterizing different subsystems and policymaking fields associated with a same dynamic and complex system;
6. Relevance of intangibles affecting the system's outcomes;
7. Nonlinear relations between causes and effects;
8. Number of policy trade-offs in time and space; and
9. Unpredictability of external factors (in respect to policy makers) that may affect the system's outcomes.

The two examples of coordinated performance in public policies discussed in the previous section can be helpful to sketch an outline of the challenges that the design and use of dynamic performance management systems should face to deal with social outcomes in different environments.

As discussed earlier, the example of health and social services mainly implies the challenge to design performance management systems that are primarily focused on the ability of different policy makers to deliver value to the benefit of service users. The dynamic and complex issues related to such context are mostly associated with the capability of different stakeholders to streamline their own policy implementation actions, so to consistently affect the desired outcomes. On the other hand, the example of food policy first implies the need to design performance management systems that may foster policy integration, leading to

consistent service delivery. This requires some previous attempts to ensure performance within each of the component organizations in the chain of production of food safety.

As shown in Fig. 8.2, the use of a dynamic performance management approach to foster policy integration and service delivery in highly dynamic and complex systems implies that different policy makers affect the system’s outcomes. Without an alignment in policy design by such players, the outcomes generated by service delivery could be poorer than the expectations of both the users and policy makers; this would reflect into inefficiency and lack of generated public value. In such a perspective, coordination and integration are central to performance, rather than being a desirable feature that might be achieved later in the policy process.

The figure also shows that service delivery outcomes (a “lower layer” group of end-results) are indirectly affected—through the performance drivers and strategic resources—by policy integration outcomes (a “higher layer” group of end-results). For instance, the number of shared decisions by different institutions involved in the value chain leading to final service delivery (a “higher layer” end-result) can be increased through the design and implementation of an integrated set of policies aimed to change a “blend” of strategic resources, such as: (1) regulations and procedures with which institutions must comply, (2) incentives and rewards systems, perhaps more inclined to reward collaborative rather than individualistic

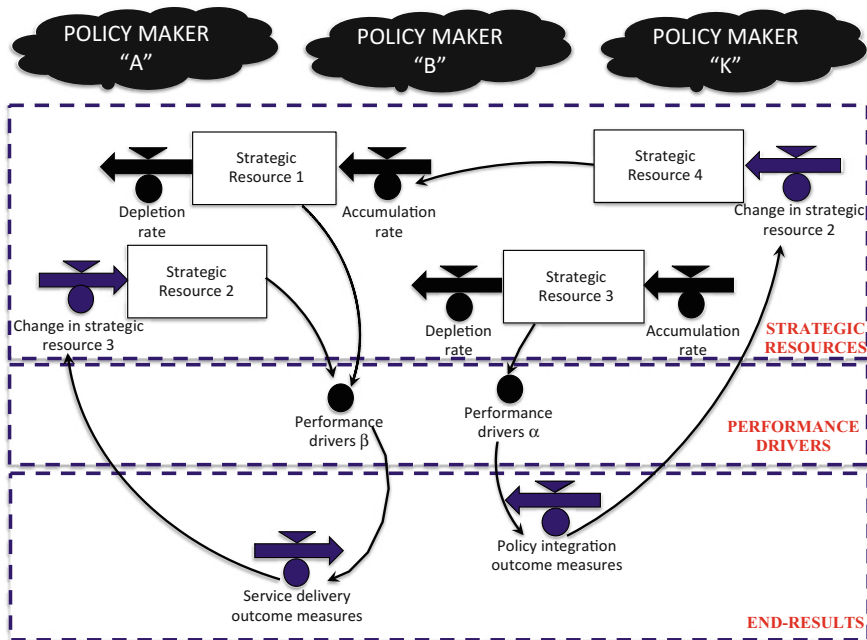


Fig. 8.2 Applying dynamic performance management to foster policy integration and service delivery in highly dynamic and complex systems: a generic model

approaches in policy making and information sharing; (3) performance measurement systems, perhaps both shared and intersectoral performance measures, oriented not only to the short but also to the long term; (4) information systems, which could be more transparent for all the involved policy makers, regardless departmental or institutional barriers; (5) cultural systems, which should gradually replace a static, sectoral, and “rule compliance” view of public administration, with a dynamic, collaborative, and outcome-based view.

Likewise, “lower layer” end-results, associated with service delivery outcomes—such as the contribution of agriculture or other industries to the gross domestic product and to the employment rates, or the environmental pollution rate—might be affected through the exploitation of the described policy integration outcomes. Such efforts might imply the design and implementation of further (lower level) policies aimed at developing other strategic resources such as: communication between public sector policy makers and other stakeholders in the system, to generate new “public–private” collaboration initiatives (an output end-result). Such end-result would accumulate into a stock (strategic resource) of total “public–private” collaborations, which would further contribute (through performance drivers) to service delivery performance outcomes.

Figure 8.3 sketches a generic dynamic performance management model related to food policy coordination. In particular, it demonstrates the expected roles of two performance drivers, i.e., those related to the quality and intensiveness of public/private sector collaborations and communication efforts. Such performance measures contribute to building up the stock of total collaboration among the relevant organization, which in turn may affect the investigated outcomes (changes in pollution and employment rates).

A similar approach could be used to model the performance factors affecting the outcomes related to social and health policies. As said, in this case, as in the previous one, the focus of the performance management model would be more on service delivery than on policy integration. However, similarly to what has been described regarding the public/private sector collaborations, also in this case

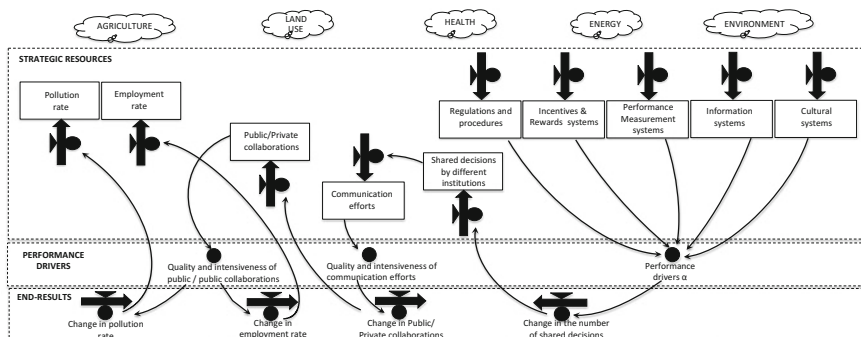


Fig. 8.3 Applying dynamic performance management to foster policy integration and service delivery: a generic model on food policy

participative governance initiatives might affect outcome measures such as the hospital readmission rates (Santos et al. 2002). Likewise, improved and shared information systems—as well as the use of rewards systems more inclined to foster collaborative approaches—would be likely to improve referral rates across agencies (outcome performance measures). Again, coordination and integration become components of service delivery rather than a separate part of policymaking.

8.5 Summary and conclusions

In summary, this chapter will make some initial steps in conceptualizing performance measurements for coordination and demonstrate how they be applied in two policy areas with different requirements for coordination. To this end, a dynamic performance management approach is suggested, and such perspective is used to frame the two investigated policy areas. It also begins to link processes of policymaking, and drivers of performance, with the final outcomes of policymaking. While not actually measuring that coordination in actual policymaking situations, it will provide the foundation for further development of this important dimension of performance management.

This approach to performance management integrates coordination more closely with performance than do more conventional performance management systems. In particular, it attempts to bring together some approaches to measuring coordination with the performance management systems. Further, the more inclusive conception of performance, both across time and across actors, to some extent returns performance management to a broader view of policy evaluation that has been lost in the emphasis on short-term measurements.

References

- Alford, J., Hartley, J., Yates, S., Hughes, O. (2016). Into the purple zone: Deconstructing the politics/administration distinction. *American Review of Public Administration*, 1–17 (online first publication, 24 March).
- Bianchi, C. (2015). Enhancing joined-up government and outcome-based performance management through system dynamics modeling to deal with wicked problems: the case of societal aging. *Systems Research and Behavioral Science*, 32, 502–505.
- Bianchi, C. (2016). *Dynamic performance management*. Berlin: Springer.
- Bianchi, C., & Rivenbark, W. C. (2013). A comparative analysis of performance management systems. *Public Performance and Management Review*, 35, 509–526.
- Boateng, H., & Agyemang, F. G. (2016). A qualitative insight into the key determinants of knowledge-sharing in a public sector institution in Ghana. *Information Development*, 17, 35–43.
- Bogdanor, V. (2005). *Joined-up government*. Oxford: Oxford University Press.
- Boston, J., & Pallott, J. (1997). Linking strategy and performance: Developments in the New Zealand public sector. *Journal of Policy Analysis and Management*, 16(3), 382–404.

- Bouckaert, G., Peters, B. G., & Verhoest, K. (2010). *The coordination of public sector organizations: Shifting patterns of public management*. Basingstoke: Macmillan.
- Bouckaert, G., & Halligan, J. A. (2008). *Managing performance: international comparisons*. London: Routledge.
- Braun, D. (2008). Organizing the political coordination of knowledge and innovation policies. *Science and Public Policy*, 35, 227–239.
- Briassoulis, H. (2005). *Policy integration and complex environmental problems: the example of Mediterranean desertification*. Aldershot, UK: Ashgate.
- Challis, L., et al. (1988). *Joint approaches to social policy: rationality and practice*. Cambridge: Cambridge University Press.
- Christensen, T., Femreite, A. L., & Laegreid, P. (2014). Joined-up government for welfare administration reform in Norway. *Public Organization Review*, 14, 439–456.
- Cook, K. S., & Cheshire, C. (2015). Social Exchange, Power and Inequality in Networks. In R. Wittek, T. A. N. Snijders, & V. Nee (Eds.), *The handbook of rational choice social research*. Stanford, CA: Stanford University Press.
- Dahlström, C., Peters, B. G., & Pierre, J. (2010). *Steering from the centre: strengthening political control in Western Democracies*. Toronto: University of Toronto Press.
- Davis, G. F., & Cobb, J. A. (2010). Resource dependency theory: past and future. In C. B. Schoonhoven & F. Dobbin (Eds.), *Stanford's Organization Theory Renaissance* (pp. 1970–2000). Manchester: Emerald Press.
- Elliot, R. (1998). Professional developments. *Exploring the purple zone, public administration and development*, 18, 81–84.
- Ghaffarzadegan, N., Lyneis, J., & Richardson, G. P. (2011). How small system dynamics models can help the public policy process. *System Dynamics Review*, 27(1), 22–44.
- Grochowska, R. (2014). Specificity of food safety concept as a wicked problem. *Journal of Agricultural Science and Technology*, 4, 823–831.
- Head, B., & Alford, J. (2013). Wicked problems: implications for public policy and management. *Administration and Society*. doi:10.1177/0095399713481601. Published online.
- Howlett, M. (2000). Managing the “Hollow State”: procedural policy instruments and modern governance. *Canadian Public Administration*, 43, 412–431.
- Husted, K., & Michailova, S. (2002). Diagnosing and fighting knowledge-sharing hostility. *Organizational Dynamics*, 31, 60–73.
- Jennings, E. T., & Crane, D. (1994). Coordination and welfare reform: the quest for the philosopher's stone. *Public Administration Review*, 54, 341–348.
- Jennings, E. T., & Ewalt, J. A. G. (1998). Interorganizational coordination, administrative consolidation, and policy performance. *Public Administration Review*, 58, 417–428.
- Jordan, A., & Lenschow, A. (2010). Environmental policy integration: a state of the art review. *Environmental Policy and Governance*, 20, 147–158.
- Laegreid, P., Rykkja, L. H. (2014). Governance for complexity—how to organize for the handling of «wicked issues»? *Stein Rokkan Centre for Social Studies*. <https://bora.uib.no/handle/1956/9384>.
- Leichsenring, K. (2004) Developing integrated health and social care services for older persons in Europe. *Integrated Journal of Health Care*, 4. <http://www.ijic.org/articles/107/>.
- Locke, A., & Henley, G. (2016). The implications of land grabs and biofuels expansion on food and nutrition security in developing countries. In B. Pritchard, R. Ortiz, & M. Shekar (Eds.), *Routledge handbook of food and nutrition security*. London: Routledge.
- Main, E. C., Bowman, R. L., & Peters, B. G. (1972). Model cities workers' perceptions of clients. *Urban Affairs Review*, 7, 309–323.
- MacRae, R. (2011). A joined-up food policy for Canada. *Journal of Hunger and Environmental Nutrition*, 6, 424–457.
- Marinotto, M. (2011). A Lipskian analysis of child protection failures from Victoria Climbié to ‘Baby P’: A street-level re-evaluation of joined-up governance. *Public Administration*, 89, 1164–1181.

- Matheson, A., Scanlan, G., Tanner, R. (1997), *Strategic management in government: Extending the reform model in New Zealand*. Wellington: State Service Commission. <http://www.oecd.org/governance/budgeting/1902913.pdf>.
- Metcalfe, L. (1994). International policy co-ordination and public management reform. *International Review of Administrative Sciences*, 60, 271–290.
- Peters, B. G. (2015). *Pursuing horizontal management: The politics of coordination*. Lawrence: University Press of Kansas.
- Peters, B. G. (2016) Bureaucratic capacity and food security in Africa, *Paper presented at conference on food security in Africa*. University of Wageningen, The Netherlands
- Pollitt, C. (2003). Joined-up government: A survey. *Political Studies Review*, 1, 34–49.
- Pritchard, B. (2016). Food and nutrition security: future priorities for research and policy. In B. Pritchard, R. Ortiz, & M. Shekar (Eds.), *Routledge handbook of food and nutrition security*. London: Routledge.
- Rittel, H. W., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, 4, 155–169.
- Santos, S. P., Belton, V., & Howick, S. (2002). Adding value to performance measurement by using system dynamics and multicriteria analysis. *International Journal of Operations and Production Management*, 22(11), 1246–1272.
- Scharpf, F. W. (1997). *Games real actors play; actor centered institutionalism in policy research*. Boulder, CO: Westview.
- Schout, A., & Jordan, A. (2005). Coordinated European Governance: Self-organizing or centrally steered? *Public Administration*, 83, 201–220.
- Shergold, P. (1997). The colour purple: perceptions of accountability across the Tasman. *Public Administration and Development*, 17, 293–306.
- Svara, J. (2001). The Myth of the dichotomy: complementarity of politics and administration in the past and future of public administration. *Public Administration Review*, 61(2), 176–183.
- Willems, A., & Buelens, M. (2007). Knowledge sharing in public sector organizations: the effect of organizational characteristics on interdepartmental knowledge sharing. *Journal of Public Administration Research and Theory*, 17, 581–606.

Chapter 9

Outcome-Based Performance Management in the Public Sector: What Role for Inter-organizational ICT Networks?

Luca Brusati, Paolo Fedele, Mario Ianniello and Silvia Iacuzzi

Abstract Public sector decision-making takes place across multiple tiers of government and societal organizations. Governance arrangements have major implications on performance, especially when addressing wicked problems: dispersed authority can lead to silos mentality, thus triggering coordination problems. We explore how inter-organizational Information and communication technology (ICT) networks can improve the performance of public sector organizations responsible for wicked problems by observing three such networks implemented by the Friuli Venezia Giulia Region (Italy): a portal for tourism promotion, a tracking system for intermodal cargo networks and a database of public works designed to prevent floods and landslides. Our analysis highlights that inter-organizational ICT networks improve outcomes when interests converge, but seem to have little impact when stakeholders perceive the interplay as a zero-sum game, and to perform best

Research was conducted in the framework of and received partial funding by the Friuli Venezia Giulia Regional Operational Programme for 2007–2013 of the European Regional Development Fund. The chapter is the outcome of the joint research work of all four authors: Luca Brusati wrote paragraph 5, Paolo Fedele wrote paragraphs 1 and 2, Mario Ianniello wrote paragraphs 3 and 4.2 and Silvia Iacuzzi wrote paragraphs 4.1 and 4.3. A preliminary version of this chapter was presented in the framework of the Conference “Contemporary Issues in Local Economic Development,” held in Rijeka University (Croatia) on September 8–10, 2014 with the title “Putting Multi-level Governance to Work for Local Economic Development: The Role of Inter-organizational ICT Networks.”

L. Brusati (✉) · P. Fedele · M. Ianniello · S. Iacuzzi
Department of Economics and Statistics, Udine University, Udine, Italy
e-mail: luca.brusati@uniud.it

P. Fedele
e-mail: paolo.fedele@uniud.it

M. Ianniello
e-mail: mario.ianniello@uniud.it

S. Iacuzzi
e-mail: silvia.iacuzzi@uniud.it

when interaction patterns are closest to hierarchy. In brief, the evidence suggests that ICT cannot contend to be a substitute for a shared governance model.

Keywords Coordination · Governance · Hierarchy · Network · Performance

9.1 Introduction

The debate about network governance (Rhodes 1997; Klijn 2008) and multi-level governance (Piattoni 2010; Toonen 2010) has become central in public policy and management in the last decades. For the purposes of this chapter, “governance” is defined as the coordination of activities around collective problems by mutually dependent actors (Koppenjan and Klijn 2004). By highlighting the key role of governance, researchers and practitioners alike acknowledge that “go alone” government-centric strategies and clear-cut lines of responsibility are not the way decision-making takes place in public sector organizations (Torfing et al. 2012; Torfing and Triantafillou 2013).

Governance arrangements have consequences at the micro-organizational level, i.e. for institutions and officials in charge for programme and project management. The dispersion of authority implies the entrenchment of “silos mentality” (Page 2005), which naturally triggers coordination problems. Coordination becomes especially critical when dealing with wicked problems, i.e. complex, intractable issues which are never wholly solved or reconciled because of the difficulty to address them, as a consequence of multiple ways of defining the nature of the problem, a lack of definitive solutions with no clear criteria for determining when the problem is solved and the potential for further problems to emerge when attempts are made to tackle them (Roberts 2000). On this backdrop, a relatively new trend such as the governance paradigm can be conceptualized as an evergreen in public management studies (Bogdanor 2005): how to enhance coordination, and thus improve performance, in the presence of wicked problems in a multi-stakeholder environment (Pressman and Wildavsky 1984).

This chapter explores whether inter-organizational Information and communication technology (ICT) networks can improve performance in the specific realm of local economic development through the analysis of a multiple case study from the Friuli Venezia Giulia (FVG) Autonomous Region in Italy. Section 9.2 reviews the relevant literature on governance as a coordination problem and the instruments used to improve system-wide performance, with a special focus on inter-organizational ICT networks. In Sect. 9.3, we outline the methodology used for our multiple-case empirical investigation. The fourth section presents the case studies, sharing relevant background information and the results of our analysis. This is followed in Sect. 9.5 by a discussion of the findings regarding coordination mechanisms in a multi-level governance setting, meant to assess whether inter-organizational ICT networks can actually improve coordination, and hence the performance of governmental intervention. The conclusions gauge the potential of

inter-organizational ICT networks as a component of the toolbox for outcome-based performance management and discuss limitations and opportunities for further research.

9.2 Literature Review

This chapter focuses on governance in one specific area of governmental intervention, i.e. local economic development. This is a typical wicked problem in which governance strongly matters, both in positive-descriptive and normative terms (Torfing et al. 2012): a large number of actors can play a role in shaping the speed and trajectories of local economic development, including local authorities, regional bodies, national and supranational tiers of governments, private firms, societal organizations. Because of the interdependence of their interests, improved outcomes can only stem from their cooperation; on the other hand, because of multiple interests and trade-offs, coordination is far from simple (Marin 1990).

As a consequence, governmental authorities in charge for local economic development face the challenge of overcoming the silos mentality and resistance that prevail across multi-stakeholder arenas in order to favour positive-sum outcomes. This is especially complex when top-down “command and control” approaches are considered incompatible with the subsidiarity principle, and the leverage of governmental expenditure cannot be used to iron out conflicting, if not opposing, interests.

The literature suggests a series of corrective actions that may result in a stronger emphasis on coordination. Coordination is a classical term in the organizational vocabulary (Groth 1999), where it means “integrating or linking together different parts of an organization to accomplish a collective set of tasks” (Van de Ven et al. 1976, p. 322). Coordination can also be described in terms of mechanisms: well-known sets include mutual adjustment, direct supervision, standardization of skills and knowledge, standardization of work processes, standardization of output and standardization of norms (Mintzberg 1997; Melin and Axelsson 2005). Communication and information are generally included in mutual adjustment mechanisms.

For the purposes of this chapter, inter-organizational coordination mechanisms in a public sector context are understood as “the instruments and mechanisms that aim to enhance the voluntary or forced alignment of tasks and efforts of organizations within the public sector. These are used in order to create a greater coherence and to reduce redundancy, lacunae and contradictions within and between policies, implementation or management” (Verhoest et al. 2007, p. 330). Redundancy occurs when two or more organizations perform the same task; lacunae when no organization performs an important task; and contradictions when the same tasks have different goals and requirements between different actors (Peters 1998).

A classification of mechanisms referring to hierarchy, market and network as coordination modes has been developed by Bouckaert et al. (2010) using the

conceptual inputs provided by Ouchi (1980), Thompson et al. (1991) and Peters (1998). Mechanisms can be either managerial (e.g. procedural rules, planning systems or input–output-oriented financial management systems) or structural (e.g. reshuffling of competences or lines of control and accountability, establishment of new bodies or structures specifically designed for the improvement of coordination). Verhoest et al. (2007) developed a typology of coordination tools and linked them to the basic coordination modes. In sum, the alternatives available in the public sector are:

- Coordination by hierarchy-type mechanisms (HTMs), based on authority and dominance. HTMs involve goal- and rule-setting, allocation of tasks and responsibilities and lines of direct control and accountability. They may also consist of traditional input-oriented financial management systems or top-down and unilateral strategic management procedures;
- Coordination by network-type mechanisms (NTMs), based on interdependence and mutual trust between stakeholders. NTMs seek common values, knowledge and strategies. Governments may create, facilitate or sustain network-like structures, such as shared information systems, collective decision-making structures or even partnership organizations. Moreover, mechanisms like staff development schemes, job rotation between public sector organizations to create a common civil service culture and other inter-organizational learning instruments like knowledge management may foster common understanding and values (Hood 2005, pp. 31–32), which are key to coordination by NTMs;
- Coordination by market-type mechanisms (MTMs), based on competition and exchange between stakeholders. MTMs aim to create performance incentives, such as result-oriented financial management systems for budgeting, accounting and audit, or incentives for agencies, units or even individual staff members. Governments may purposefully create and safeguard markets (e.g. internal and quasi-markets) to foster coordination by encouraging intra- and inter-organizational competition.

Based on this menu of coordination modes, this chapter explores the potential of inter-organizational ICT networks as a solution that promises to facilitate both horizontal and vertical coordinations, and hence improve system-wide performance *vis-à-vis* a wicked problem such as local economic development.

At the level of individual organizations, ICT networks provide a tool to store and retrieve information, accomplish tasks and impose a schedule on processes given their ability to perform and memorize actions as well as to permit, promote and facilitate the performance of actions by users, both through the information system itself and based on information from a system (Orlikowski 1991; Malone and Crowston 1994). A broad literature exists addressing the consequences of adopting ICTs, or “e-government,” at the level of individual public sector organizations (e.g. Brown 2005; Beynon-Davies and Martin 2010; Weerakkody and Reddick 2013), but the relationship between ICTs and network governance has been studied almost exclusively by political scientists interested in the possibility to harness ICT

network to facilitate citizens' participation in decision-making (e.g. Snellen 2007; Misuraca et al. 2011; Adesola 2012; for an exception see Brown et al. 1998).¹

Inter-organizational ICT networks include the full range of telecommunications networks, information technologies (IT) and electronic services (e-services) which may be set up to facilitate coordination across organizations. ICTs are important tools to support public sector governance since they facilitate learning by sharing information and solutions among different levels of government and society. The term "ICT networks" would seem to imply that their underlying logic is network-type, i.e. that they are predominantly NTMs; yet, as the cases analysed in this chapter show, the design and implementation of ICT networks can feature hierarchical or market-type mechanisms as well.

9.3 Method

Following the approach to case study research proposed by Eisenhardt (1989), Stake (2005) and Yin (2014), and more specifically by Stewart (Stewart 2012) with reference to governance research, this chapter explores the role of inter-organizational ICT networks in promoting local economic development by analysing the results of a multiple case study concerning three targeted interventions implemented by the FVG Regional Administration. Regions are the intermediate tier of government in the Italian system of public administration: they are not primarily responsible for direct service delivery, but rather for planning and policy coordination. As noted by several studies, regional and municipal decision-makers are closer to the sources of innovation than those at national level (Kirchherr et al. 2014) and are therefore better placed to support economic development. For historical reasons, five so-called "autonomous" regions in Italy enjoy enhanced coordination powers, in particular in the field of local economic development; out of them, FVG is the median region in terms of population size, and for this reason it was selected as a suitable empirical locus for our analysis.

Since 2009, the FVG Regional Administration promoted the establishment of inter-organizational ICT networks as a way to enhance the effectiveness of local economic development. Our analysis focused on three interventions meant to promote tourism, transportation and soil conservation respectively, i.e.,

1. An integrated portal for tourism promotion, doubling up as extranet for electronic data interchange (EDI);
2. An integrated tracking system for cargo moving across intermodal networks;
3. An integrated database of all public works designed to prevent floods and landslides.

¹A burgeoning stream of literature exists on "ICT for development," but the focus of these studies is on the potential for ICT to trigger macro-economic development in low-income countries (Avgerou 2010); as such, they fall beyond the scope of this chapter.

The interventions analysed in our fieldwork share by design important commonalities in size, scope and time frame, since they were all envisaged as part of the FVG Regional Operational Programme of the European Regional Development Fund for the 2007–2013 budgeting period (*Programma Operativo Regionale del Fondo Europeo di Sviluppo Regionale*, or POR FESR 2007–2013): this is the EU-cofinanced planning tool with which European regions are encouraged to kick-start projects designed to support the development of their own economies. EU-cofinanced interventions lend themselves well to comparative analysis, since they have a beginning and an end date, and are designed, implemented and evaluated according to standardized procedures.

The bodies in charge for the programme funded the establishment of inter-organizational ICT networks as a way to enhance regional competitiveness by reducing the fragmentation of communication and information flows and enhancing the accessibility and usability of regional structures. The interventions selected for our study belong to two different strategic priorities of POR FESR 2007–2013: one was funded in the framework of Axis 2, i.e. environmental sustainability, which focused on caring for existing resources as well as avoiding adverse environmental impacts; two more were funded in the framework of Axis 3, i.e. accessibility, which pursued better integration of the regional economy by developing transportation and information services. We selected purposefully interventions in different sectors so as to be able to compare the impact on performance of inter-organizational ICT networks established in areas of governmental interventions characterized by different governance challenges.

Data were gathered through a triangulation of documental analysis, direct observation and semi-structured interviews with key informants. ICT networks were not studied in terms of their technical features, but rather focusing on their capacity of enhancing inter-organizational coordination. Section 9.4 describes each case study separately, then in Sect. 9.5 comparisons are drawn to better understand the role that inter-organizational ICT networks can play in promoting system-wide performance.

9.4 Empirical Data and Analysis

9.4.1 “Regional Tourism Information System”: *An Integrated Portal for Tourism Promotion*²

The large number and diversity of public and private stakeholders that characterize the tourism sector in FVG as elsewhere, including regional promotion offices, municipalities, consortia, private tourism agencies, accommodation facilities,

²“Sistema informativo turistico regionale FVG”; POR FESR 2007–2013, Asse 3, Attività 3.2.b “Sviluppo servizi informatici avanzati per il sistema turismo” (Development of advanced IT services for the tourism system).

restaurants, entertainment and cultural outfits, led to significant fragmentation of information and communication flows, both towards incoming tourists as well as within the sector. Fragmentation triggers conflicts and inefficiencies, which increase costs and make it difficult to target promotional messages to the right market segments, thus hindering the effective promotion of regional tourism.

The overall project objective was to help increase the number of tourists visiting FVG by establishing a completely revamped web portal meant to serve as the leading point of reference for both tourists and the hospitality industry. The aim was to increase the quality and quantity of information available at once to incoming tourists and all other stakeholders, while enabling an integrated monitoring of information flows. On the one hand, the new portal with a single platform shared by all stakeholders would allow for an interactive presentation of the entire offerings of FVG to incoming tourists, as well as showcase the programmes of individual tourist agencies and the deals of individual hospitality businesses. On the other hand, the portal would facilitate information exchange among stakeholders and optimize administrative processes such as statistical data collection on visitors and occupancy rates, as well as price monitoring by regional authorities. It was expected that this ICT system would allow for an integrated marketing of the entire regional tourism sector, which in turn would enhance the attractiveness, and hence the competitiveness, of hospitality businesses in FVG *vis-à-vis* both national and international markets.

The main project partners were regional agencies and private companies responsible for implementing the IT solutions:

- The Regional Service for Tourism Promotion, Internationalization and Development within the Central Directorate for Production, Commerce, Cooperation, Agriculture and Forestry, in charge for outlining the development framework for the tourism sector;
- TurismoFVG, the then Regional Tourism Agency, in charge for overall project coordination;
- Insiel, the regional in-house IT provider, in charge for establishing the revamped portal, with all annexes and integrations;
- Icon, a private firm which provided the content management system (CMS) platform;
- Umana and Obiettivo Lavoro, two human resource agencies that selected the external staff to be involved in the project.

TurismoFVG involved in consultations the representatives of all stakeholders, and in particular Insiel, responsible for setting up the new portal. Insiel created and implemented four distinct pieces of software:

- a revamped portal in Italian, English and German (<http://www.turismofvg.it>), which allows each firm in the sector (tourist agency, hotel, camping site, etc.) to create dedicated contents using a free-of-charge interface hosting a description of the various features of their facility, a management system (CRM, ERP or

similar) and a booking system; each entry can then be linked to the web sites of municipalities, tourist consortia and the like;

- an intranet which connects all stakeholders of the regional tourism system, both private and public, and which has become a repository for documents, a management tool for bookings and a support system for various administrative procedures;
- a statistical system for collecting data on incoming tourists, available facilities, etc. as well as monitoring prices (when the project was designed all accommodation facilities had to inform by law regional authorities of the prices they charged for standard services such as single and double rooms during low, middle or high season);
- a photo gallery management system which could help all those businesses which are not yet equipped with their own website or with such a feature.

Once the ICT system was ready, TurismoFVG organized several workshops for hoteliers, hospitality managers, municipalities, tourism promotion consortia, travel agencies and the like. Workshops were meant to provide tourism operators with the knowledge and tools needed to develop and implement an effective Internet marketing strategy using the regional portal: participants learnt how to use the online booking system, share information on social networks and prepare customized brochures to help tourists plan their stay using the resources available on the portal. The indirect objective of the workshops was to share multi-channel marketing solutions in line with TurismoFVG's strategies.

The coordination tools used in this case study were therefore both management (the initial consultation and design phase) and structural instruments (the actual ICT systems). While the initial phase was both inclusive and dominated by regional agencies, and as such a combination of HTMs and NTMs, the ICT platforms and subsequent training sessions were a mixture of structural and management instruments, all predominantly NTMs in nature. In the second phase, the only element of hierarchical accountability was the obligation for hospitality businesses to inform on a semi-annual basis the Regional Service for Tourism Promotion, Internationalization and Development of their prices through the intranet within the portal. However, even this requirement disappeared due to changes in the regional legislation on tourism, as hospitality businesses are now required to inform municipalities rather than regional authorities. On the other hand, the indirect economic incentive in obtaining free access to a portal, an intranet, a booking and data management system and a photo gallery can be interpreted as a structural MTM. Table 9.1 summarizes the nature of coordination tools used in case study #1.

Table 9.1 Coordination tools used in case study #1

Instruments	Mechanisms		
	HTMs	NTMs	MTMs
Management	Only initially	Yes	–
Structural	Only initially	Yes	Yes (indirectly)

It is too early to assess whether the project has reached all of its objectives, but the regional tourism system surely increased its visibility, with the average daily number of visits quadrupling from 2500 to 10,000 once the revamped regional portal went live. This steep increase shows how the new ICT network follows a market approach, since it meets the information needs of end customers, i.e. tourists, and this should increase regional competitiveness. Moreover, the new system guaranteed better coordination of the promotional strategies in the region and a rationalization of administrative procedures, while allowing for flexibility and customization by individual businesses. The main weakness of the project is precisely that very few private firms, compared to expectations, exploited the potential of the new portal by creating dedicated contents. Even though the section related to packages is up to date, an indication of how little the portal is used by businesses is that the last question present in the forum was entered in November 2013. In this case, the ICT solution did manage to decrease fragmentation, but not to overcome silos mentality. Neither network mechanisms nor the indirect market incentive of reduced costs (free access to the portal and increased visibility by linking one's entry to key sites) seem enough of an appeal for most businesses.

9.4.2 “Safe and Efficient Cargo”: An Integrated Tracking System for Containers Moving Across Intermodal Networks³

The ports and intermodal freight network infrastructure in FVG were modernized in recent years, but their potential has been weakened by their poor integration. This is due partly to the heterogeneity of the stakeholders involved and the high degree of institutional hierarchy, but it is also a consequence of the limited application of IT solutions. This entails little efficiency in entry and exit operations, with complex procedures and long waiting times for loading and offloading. Several studies point out that the competitiveness of Adriatic ports, including those in FVG, depends on their ability to enhance intermodality through the integration of land and sea transport, as well as pay attention to environmental concerns (EUNETMAR 2013). The poor integration of the regional logistic centres was therefore identified as a major hindrance to the attractiveness of FVG, as using its logistic facilities often resulted in delays and additional costs for carriers. Besides efficiency, security was also an issue: EU directives on the traceability of dangerous goods became much more stringent following the terrorist attacks of 11 September, 2001, and the FVG network did not always comply with the new regulations.

³“Safe and Efficient Cargo”; POR FESR 2007–2013, Asse 3, Attività 3.1.b “Interventi immateriali nell’ambito delle infrastrutture di trasporto” (Intangible interventions in the field of transportation infrastructure).

The overall project objective was, therefore, to help increase the volume of cargo transported through FVG with a more competitive transport and logistic system, thanks to a better integrated and streamlined system for monitoring the flow of containers by sea and land through the digitalization of shipping documentation and the tracking of vehicles carrying dangerous goods. Once fully operational, the project was meant to monitor at least 80% of the goods in transit in the regional cargo hubs for both security and traceability purposes.

The main project partners were:

- The Regional Mobility Service within the Central Directorate for Infrastructure, Mobility, Planning and Public Works, in charge for outlining the development framework for the logistic sector;
- Insiel, the regional in-house IT provider, in charge for overall project coordination;
- Five regional cargo hubs (i.e. ports and freight centres), namely Trieste Port Authority, Monfalcone Port Special Agency, the Alpe Adria freight exchange centre in Cervignano del Friuli, SDAG Gorizia and the Ferneti Terminal in Trieste, responsible for using identification and monitoring infrastructures upon project completion.

The project envisaged a full-fledged feasibility study as its first step, with technical workshops involving all relevant stakeholders. In the second phase, the new monitoring systems had to be planned, installed and made operational; for this purpose, terms of reference were prepared, proposals were received and evaluated, and technical partners were selected. The operational part then focused on implementing and testing the new system, which features four components:

- An IT system for the identification and management of goods and people accessing regional cargo hubs; this entailed laying optical fibre cables and installing detection systems with related access barriers;
- An IT system for monitoring dangerous goods; once the system is operational, containers with dangerous cargo will be automatically identified and reported to the management centres of the road operators in FVG, be they public (ANAS in Padriciano) or private (Autovie Venete in Palmanova);
- A digital system for the dematerialization of documents accompanying goods in transit;
- “Logistica FVG,” an integrated portal and control centre managed by the Regional Mobility Service within the Central Directorate for Infrastructure, Mobility, Planning and Public Works; the portal includes all statistical data collected through the other IT systems in order to improve processes and planning based on empirical evidence.

According to the classification developed by Bouckaert et al. (2010), this case study used initially management tools and then more structural ones. The feasibility study is an example of a management NTM, while the second operational phase is structural but mainly hierarchical (HTM), both in the establishment of the ICT

Table 9.2 Coordination tools used in case study #2

Instruments	Mechanisms		
	HTMs	NTMs	MTMs
Management	Yes	Only initially	–
Structural	Yes	–	–

network and its management by freight centre authorities and the Regional Mobility Service (Table 9.2).

The project is considered successful, notwithstanding some delays in implementation due to partners' heterogeneity and organizational complexity, as well as the fragmentation of decision-making responsibilities among cargo hubs, regional authorities and the national government. The project enhanced the effectiveness of past infrastructural investments: it is expected to result in the strengthening of intermodal logistic services and of their promotion through an integrated portal. This focus highlights awareness of the fact that carriers choose an infrastructure over another not only in terms of its technical features, but also in terms of cost, quality and effectiveness (The European House—Ambrosetti 2013). The improved coordination of all stakeholders has also allowed for increased service customization, and therefore client satisfaction, for example, by solving specific problems through dedicated software solutions. Last but not least, the project tackled successfully dead times and lengthy procedures: the reduction of entry, transit and exit times helps avoid congestion and thus reduces both operators' costs and environmental impact. In this case, therefore, it seems fair to conclude that inter-organizational ICT networks contributed favourably to effective local economic development.

9.4.3 “Survey System for Soil Conservation”: An Integrated Database of All Public Works Designed to Prevent Floods and Landslides⁴

Since the 1990s, the Regional Service for the Management for Rural Areas and Irrigation has established a registry for hydro-geologically instable areas and for conservation and stabilization public works carried out over the years. Such registry turned over the years into a database with about 35,000 entries; it was then merged with the *Sistema informativo geografico difesa del suolo* (Geographic information system for soil conservation), and it is now managed by the Regional Department for Civil

⁴“Sistema rilievi difesa del suolo - Aggiornamento del catasto delle opere di difesa del suolo con nuove procedure e tecnologie di acquisizione dati (opere idraulico-forestali)”, i.e. “Survey system for soil conservation - Updating of the register of the soil conservation works with new procedures and data acquisition technologies (hydraulic-forestry works)”; POR FESR 2007–2013, Asse 2, Attività 2.1.c “Prevenzione e gestione dei rischi” (Risk prevention and management).

Protection, which has made it available to all the institutional stakeholders involved in agricultural and rural planning, as well as in flood and landslide prevention.

The multiplicity of players involved in updating the database carries the risk of data inconsistency, especially if information is not collected and recorded according to the same criteria and procedures. The “Survey system for soil conservation” project aimed at harmonizing both existing entries and any future data collected, to ensure proper updating and thus the usability and reliability of the database, especially vital in case they must be used during an emergency. The main project partners were regional offices and agencies such as the Regional Service for the Management for Rural Areas and Irrigation, the Forest Management and Wood Production Service, the Forest Service and the Department for Civil Protection. Municipalities are also involved, as the stakeholders often directly involved in the design and maintenance of conservation and stabilization public works.

The project was completed in June 2014, following the implementation of four components. The first component focused on the procurement of nine workstations (“survey kits”) for proper data acquisition using latest technologies such as global positioning system (GPS) receivers and laser rangefinders with customized software for data collection and database synchronization. These survey instruments were used for both updating the existing database and classifying new entries. The second component allowed to apply radio frequency identification (RFID) technology to all conservation and stabilization public works included in the database: such microchips have a RFID tag or a transponder that allows each one of them to be uniquely recognized with a reader. For the third component, an unmanned aerial vehicle (commonly known as drone) was purchased for surveying difficult access areas or dangerous sites. The drone is controlled by two radios: the former has the flight controls, while the latter controls the sensor, which according to the needs can be a high-resolution camera, a video camera or an infrared camera, for example, for monitoring wildfires. The fourth component addressed capacity building: once the procedures for the procurement of all hardware and software components were completed, training sessions were organized in order to endow not only surveyors, but also inspectors and other operators with the knowledge and skills required to use effectively the new equipment and the renewed database. The overall guiding principle was to create an integrated, self-feeding regional information system, using flexible and high-performance technologies with low operating costs. The system is expected to facilitate the coordination of all institutional stakeholders involved in soil conservation, and thus the prevention and management of natural disasters such as floods, land erosion and landslides.

As far as the classification developed by Bouckaert et al. (2010) is concerned, differently from the other two cases, the “Survey system for soil conservation” project saw first the application of structural coordination mechanisms, and then of management ones. The updated database is, once more, an example of a structural NTM shared by all stakeholders. The new technologies and equipment were instead purchased and introduced by regional agencies, using predominantly hierarchical structural and management tools (HTM). On the other hand, the training sessions open to all potential operators, not only surveyors, are an example of management

Table 9.3 Coordination tools used in case study #3

Instruments	Mechanisms		
	HTMs	NTMs	MTMs
Management	Yes	Yes	–
Structural	Yes	Yes	–

NTM as they aim at enhancing coordination through common knowledge and values (Table 9.3).

This case study demonstrates how the implementation of inter-organizational ICT networks can contribute to outcome-based performance management by improving governance and public services delivery. Not only the project has made data acquisition more reliable, but also it has made surveyors’ work easier, safer and less time-consuming. As an additional advantage, the same tools could be used for other types of monitoring, such as those concerning crops and flora, wildlife, environmental compliance, wildfires and other natural disasters. On the other hand, though, many operators are experiencing difficulties with the new devices and procedures: many already work extra hours and have little time to invest in training. Moreover, the average surveyor’s age is rather high as a consequence of the hiring freezes that hit periodically Italian public administration: resistance to drop entrenched organizational practices makes it more complex, at least for some, to learn and use regularly new procedures, especially when they involve new technologies. Once more, ICT networks offer the opportunity to reduce fragmentation in a cost-effective manner, but can do little by themselves against inter-organizational silos mentality.

9.5 Results and Discussion

The projects “Regional tourism information system,” “Safe and efficient cargo” and “Survey system for soil conservation” are all examples of interventions designed to promote local economic development through the improved coordination made possible by inter-organizational ICT networks. The fragmentation of stakeholders and competencies was impacting adversely the competitiveness of important sectors for the FVG economy, such as tourism, logistics and soil conservation. These targeted projects are interesting examples of how it is possible to pursue improved coordination not only with top-down mechanisms, but also with bottom-up tools. All three projects involved a certain degree of hierarchy, combined with a network approach. Decision-making, planning and implementation tended to be entrusted by central bodies, i.e. regional agencies, while consultation and training involved all stakeholders. Little use was made of market-type mechanisms such as economic incentives, which are likely to be preferred when a reallocation of resources is among the goals pursued as a consequence of improved coordination. All projects, however, were implemented with an explicit focus on the need to reduce the

indirect costs stemming from poor system-level coordination, which is particularly important to deal effectively with the increasing competition characterizing both tourism and logistics.

Notwithstanding the commonalities among the projects, inter-organizational ICT networks performed differently in the three cases. In case study #1, with a prevalence of structural instruments and network-type mechanisms and limited reliance on hierarchy and market-type coordination, results were hindered by the limited success in engaging private stakeholders in the tourism sectors. Incoming tourists reacted positively to the new portal, but the new system did not attract the intended beneficiaries, i.e. the hospitality industry. In case study #2, with a prevalence of structural instruments and hierarchical coordination, the outcome seems to be more favourable, albeit obtained with delay because of the different interests to be accommodated; nevertheless, once the cargo tracking system is fully operational, it will guarantee an improved monitoring of containers moving across intermodal networks. Lastly, case study #3 presented a mixture of structural and management instruments implemented through both hierarchical and network-type mechanisms. The project was successful at updating the database and ensuring that new data can be collected with the latest technologies; not all surveyors can or will use the new data entry system, though, and this might hinder the quality, quantity and cost-effectiveness of the information available in the future.

It is too early to assess the impact of these projects on the economic development of FVG. Inter-organizational ICT networks, however, did contribute to an improved performance in two ways: on the one hand they created a shared informational background for system-level decision-making, while on the other they insulated some activities from conflict and negotiations through default mechanisms. The former outcome can be observed in all three examples; the latter was partially relevant in case study #2 and case study #3, where the cargo tracking procedures and the public works data collection system were decided and implemented in a top-down fashion by regional agencies.

At the same time, a comparison of the three cases shows that “soft” tools relying on a bottom-up logic, such as inter-organizational ICT networks, offer a facilitating condition but do not in themselves help improve outcomes in the management of wicked problems. They do help tackle some coordination issues, such as asymmetric information, but they seem to be less effective on substantial fragmentation and “silos mentality” issues in the absence of system-wide governance arrangements already in place. In other words, this multiple case study suggests that the introduction of inter-organizational ICT networks is not an alternative to the pursuit of a system-wide framework for the governance of wicked problems. Indeed, in the tourism sector, where this shared governance model is *de facto* absent, the new integrated ICT network is largely underutilized by its intended beneficiaries. The underlying driver of these adverse outcomes seems to be that the weakness of a forward-looking management culture among small and medium enterprises in the hospitality industry and a tradition of governmental subsidies lead to perceive improved coordination as a negative sum game, with extra efforts unmatched by demonstrated short-term benefits. On the other hand, in the soil conservation

system, where the regional administration is clearly the “playmaker” and there is no competition with other public sector institutions such as municipalities, an inter-organizational ICT network allows to improve outcomes for all the stakeholders involved. Lastly, the integrated tracking system for containers moving across intermodal networks works relatively well, because it involves primarily public sector institutions sharing a multi-level governance arrangement that private firms in the logistics industry must comply with.

We can conclude that, on the one hand, coordination was improved in all the three cases, and the new integrated ICT networks generate “big data” suitable to improve system-level decision-making; this is expected to lead to cost reductions and increased competitiveness, i.e. contribute positively to the overall performance of FVG in terms of local economic development. On the other hand, our multiple case study points out that silos mentality hinders the improvement of system-wide performance, also when the attempt to coordinate multiple stakeholders is carried out through inter-organizational ICT solutions.

Further research is certainly needed beyond this exploratory study, which was meant to explore the link between inter-organizational ICT networks and outcome-based performance management. It should look beyond the specific context of one Italian region in its sampling of case studies, and take into consideration other sectors and performance management tools that can help reduce fragmentation and silos mentality (Flyvbjerg 2006; Tsang 2012; Elman et al. 2016). A comparison of the three cases points out nevertheless that inter-organizational ICT networks do little to address what is, or is perceived to be, a negative or a zero-sum game, and help most when the logic underpinning inter-organizational relationships is closest to hierarchy. Ultimately, the implication stemming from this multiple case study is that inter-organizational ICT networks facilitate the management of multi-stakeholder relationships when coordination arrangements are already in place, but cannot achieve the same outcome by themselves when there is no shared governance model, and stakeholders’ interests are not aligned to begin with. In other terms, our analysis suggests that, in order to improve performance, it is not enough to invest only on facilitating conditions, such as inter-organizational ICT networks, hoping that outcomes will materialize as a sort of spillover effect: efforts must be focused on effective network governance patterns first.

References

- Adesola, S. A. (2012). Entrenching democracy and good governance: The role of ICT. *Mediterranean Journal of Social Sciences*, 3(15), 80–87.
- Avgerou, C. (2010). Discourses on ICT and development. *Information Technologies & International Development*, 6(3), 1–18.
- Beynon-Davies, P., & Martin, S. (2010). Electronic local government and the modernization agenda: Progress and prospects for public service improvement. *Local Government Studies*, 30(2), 214–229.
- Bogdanor, V. (Ed.). (2005). *Joined-up government*. Oxford: Oxford University.

- Bouckaert, G., Peters, B. G., & Verhoest, K. (2010). *The coordination of public sector organizations. Shifting patterns of public management*. Basingstoke: Palgrave Macmillan.
- Brown, D. (2005). Electronic government and public administration. *International Review of Administrative Sciences*, 71(2), 241–254.
- Brown, M. M., O’Toole, L. J., Jr., & Brudney, J. L. (1998). Implementing information technology in government: An empirical assessment of the role of local partnerships. *Journal of Public Administration Research and Theory*, 8(4), 499–525.
- Eisenhardt, K. (1989). Building theories from case study research. *Academy of Management Review*, 14(4), 532–550.
- Elman, C., Gerring, J., & Mahoney, J. (2016). Case study research: Putting the quant into the qual. *Sociological Methods Research*, 45(3), 375–391.
- EUNETMAR. (2013). Analysis to support the elaboration of the Adriatic and Ionian Maritime action plan, *Studies to Support the Development of Sea Basin Cooperation in the Mediterranean, Adriatic and Ionian, and Black Sea*, March 31st, Brussels. Retrieved December 19, 2016, from https://webgate.ec.europa.eu/maritimeforum/sites/maritimeforum/files/Report%202_21_03_2014.pdf.
- Flyvbjerg, B. (2006). Five misunderstandings about case study research. *Qualitative Inquiry*, 12(2), 219–245.
- Groth, L. (1999). *Future organizational design—The scope for the IT-based enterprise*. Chichester: Wiley.
- Hood, C. (2005). The idea of joined-up government: A historical perspective. In V. Bogdanor (Ed.), *Joined-up government*. Oxford: Oxford University Press.
- Kirchherr, J., Scherf, G., & Suder, K. (2014). *Creating growth clusters. What role for local government?* Berlin: McKinsey.
- Klijin, E. H. (2008). Governance and governance networks in Europe: An assessment of ten years of research on the theme. *Public Management Review*, 10(4), 505–525.
- Koppenjan, J. F. M., & Klijin, E. H. (2004). *Managing uncertainties in networks*. London: Routledge.
- Malone, T. W., & Crowston, K. (1994). The interdisciplinary study of coordination. *ACM Computing Surveys*, 26(1), 87–119.
- Marin, B. (1990). Generalized political exchange: Preliminary considerations. In B. Marin (Ed.), *Generalized political exchange: Antagonistic cooperation and integrated policy circuits*. Frankfurt: Campus Verlag.
- Melin, U., & Axelsson, K. (2005). Understanding organizational coordination and information systems: Mintzberg’s coordination mechanisms revisited and evaluated. In *European conference on information systems (ECIS)—Information systems in a rapidly changing economy* (pp. 1634–1644), May 26th–28th, Regensburg.
- Mintzberg, H. (1997). The strategy concept I: Five Ps for strategy. *California Management Review*, 30(1), 11–24.
- Misuraca, G., Alfano, G., & Viscusi, G. (2011). A multi-level framework for ICT-enabled governance: Assessing the non-technical dimensions of ‘Government Openness’. *Electronic Journal of e-Government*, 9(2), 152–165.
- Orlikowski, W. J. (1991). Integrated information environment or matrix of control: The contradictory implications of information technology. *Accounting, Management and Information Technology*, 1(1), 9–42.
- Ouchi, W. G. (1980). Markets, bureaucracies, and clans. *Administrative Science Quarterly*, 25(1), 129–141.
- Page, E. (2005). Joined-up government and the civil service. In V. Bogdanor (Ed.), *Joined-up government*. Oxford: Oxford University Press.
- Peters, B. G. (1998). Managing horizontal government: The politics of co-ordination. *Public Administration*, 76, 295–311.
- Piattoni, S. (2010). *The theory of multi-level governance. Conceptual, empirical, and normative challenges*. Oxford: Oxford University Press.

- Pressman, J. L., & Wildavsky, A. (1984). *Implementation* (2nd ed.). Berkeley: University of California Press.
- Rhodes, R. A. W. (1997). *Understanding governance*. Buckingham: Open University Press.
- Roberts, N. C. (2000). Wicked problems and network approaches to resolution. *International Public Management Review*, 1(1), 1–19.
- Snellen, I. (2007). E-government: A challenge for public management. In E. Ferlie, L. Lynn, & C. Pollitt (Eds.), *The Oxford handbook for public management* (pp. 398–421). Oxford: Oxford University Press.
- Stake, R. E. (2005). Qualitative case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (pp. 443–466). Thousand Oaks, CA: Sage.
- Stewart, J. (2012). Multiple-case study methods in governance-related research. *Public Management Review*, 14(1), 67–82.
- The European House—Ambrosetti. (2013). Il rilancio della portualità e della logistica italiana come leva strategica per la crescita e la competitività del Paese. In *Il rilancio della portualità e della logistica italiana come leva strategica per la crescita e la competitività del Paese*, June 18, Rome. Retrieved August 14, 2014, from <http://www.ambrosetti.eu/it/download/ricerche-e-presentazioni/2013/il-rilancio-della-portualita-e-della-logistica-italiana-come-leva-strategica-per-la-crescita-e-la-competitivita-del-paese>.
- Thompson, G., Frances, J., Levacic, R., & Mitchell, J. (1991). *Markets, hierarchies and networks: The coordination of social life*. London: Sage.
- Toonen, T. A. J. (2010). Multi-level governance and intergovernmental relations: Integrating the theoretical perspectives. In E. Ongaro, A. Massey, M. Holzer, & E. Wayenberg (Eds.), *Governance and intergovernmental relations in the European Union and the United States: Theoretical perspectives* (pp. 29–50). Cheltenham: Elgar.
- Torfin, J., & Triantafillou, P. (2013). Introduction to interactive policy making, metagovernance and democracy. In J. Torfin & P. Triantafillou (Eds.), *Interactive policymaking, metagovernance and democracy* (pp. 1–25). Oxford: Oxford University Press.
- Torfin, J., Peters, B. G., Pierre, J., & Sørensen, E. (2012). Assessing and improving effective interactive governance. In J. Torfin, B. G. Peters, J. Pierre, & E. Sørensen (Eds.), *Interactive governance. Advancing the paradigm* (pp. 166–187). Oxford: Oxford University Press.
- Tsang, E. W. K. (2012). Case study methodology: Causal explanation, contextualization, and theorizing. *Journal of International Management*, 19(2), 195–202.
- Van de Ven, A. H., Delbecq, A. L., & Koenig, R., Jr. (1976). Determinants of coordination modes within organizations. *American Sociological Review*, 41(2), 322–338.
- Verhoest, K., Bouckaert, G., & Peters, B. G. (2007). Janus-faced reorganization: Specialization and coordination in four OECD countries in the period 1980–2005. *International Review of Administrative Sciences*, 73(3), 325–348.
- Weerakkody, V., & Reddick, C. G. (Eds.). (2013). *Public sector transformation through e-government: Experiences from Europe and North America*. London: Routledge.
- Yin, R. K. (2014). *Case study research: Design and methods* (5th ed.). Thousand Oaks, CA: Sage.

Chapter 10

Disappointing Outcomes: Can Implementation Modeling Help?

I. David Wheat and Eugene Bardach

Abstract This paper addresses questions about modeling the implementation requirements of a public policy proposal. Can modeling provide advance warning of problematic implementation requirements inherent in the design of a policy idea? Going further, can it suggest feasible redesign options to improve the chances for desired outcomes? Our methodology, system dynamics, is more than just a simulation tool; it also a method of scientific inquiry that fosters operational thinking about how to improve the functioning of complex social systems. Our model is motivated by a case often cited as the seminal work in the implementation literature: Pressman and Wildavsky's narrative of problems that undercut a US policy to combat persistent unemployment among minorities in Oakland, California in the late 1960s.

Keywords Implementation · Public policy · System dynamics

The classic case of big projects having little effect is the 'Oakland' fiasco famously analyzed by Pressman and Wildavsky (1973). Their book launched the implementation research agenda for the public policy discipline, guided by the hypothesis that 'separation of policy design from implementation is fatal' (Pressman and Wildavsky 1973, xxiii). We previously used the Oakland case to illustrate the benefits of interdisciplinary collaboration between scholars in the fields of public policy and system dynamics, and this chapter builds on that earlier effort.¹

¹The first version, "Public Policy Implementation Modeling: The Case of EDA in Oakland," was presented at the International System Dynamics Society Conference in Boston in July 2015. A substantially revised version was presented at the IJPA Symposium at the University of Palermo in May 2016.

I.D. Wheat (✉)
System Dynamics, University of Bergen, Bergen, Norway
e-mail: david.wheat@uib.no

E. Bardach
Goldman School of Public Policy, University of California, Berkeley, Berkeley, USA
e-mail: ebardach@berkeley.edu

Our approach is framed by two questions related to the implementation requirements of a public policy initiative. First, can modeling reveal those implementation requirements and the potential for disappointing outcomes? And, can a series of simulations under different assumptions about a policy suggest feasible redesign options to improve the chances for desired outcomes?

In the first section, we discuss distinctive features of implementation problems in the public realm and make the argument for a system dynamics (SD) approach to some of those problems. The second section provides a brief overview of SD-based qualitative implementation modeling, using a hypothetical policy issue to illustrate the method. The third section is a more detailed examination of a quantitative approach that utilizes SD simulation methods to explore economic development projects involving government and private sector partnerships, and the Oakland case provides our illustration. Finally, we conclude with take-away messages about the value of both qualitative and quantitative implementation modeling, and suggest ways that others might integrate their methods with the approach presented here.

10.1 Implementation and Policy Design

Implementation appeared in 1973, but the seamless web of policy design and implementation was recognized long before the 1970s; e.g., Carl Friedrich observed in 1940 that the ‘formation’ of public policy ‘is inseparable from its execution’ (cited in Wilcox 1978). For much of the twentieth century, however, the dominant paradigm encouraged a research demarcation between the formulation of policy and its implementation; the former involving politics and the latter involving ‘mere’ administration. See Wheat (2010) for a brief historical review of the paradigm shift that occurred in the 1970s after the much-publicized implementation failures of some Great Society programs.

Continuous policy resistance in the public arena accounts for some of the observed gaps between public program outputs and their impact. The political conflicts that have beset the adoption process do not disappear during the implementation process; in some cases, they may be aggravated. New conflicts may appear, lured out of hiding by issues that come up during implementation but had been suppressed or invisible previously. These conflicts, together with the problems of turning a policy over to existing public-sector bureaucracies and perhaps to a host of private-sector partners at the same time, guarantee a rocky implementation process. The results, frequently, are delay, erosion of policy goals, cost overruns, the intrusion of various interests seeking to capture economic rents, and a degradation of whatever future operational capacity was envisioned.

How might policy designers cope with the contingencies and probable setbacks of the implementation process? First, they must take some responsibility for implementation and avoid assuming it is someone else's job.² They can do this by anticipating implementation issues during the design process and crafting policies that would be reasonably robust against the difficulties of implementation. This means building in extra time for delays caused by busy or uncooperative bureaucracies, budget problems caused by overly optimistic financial planning, the sacrifice of certain goals to political and administrative compromises, and workarounds that lead to building a program out of components (such as a certain proportion of untrained or incompetent personnel) that are less well suited to the task than originally assumed. In other words, at the design stage, it may be possible to anticipate potential implementation obstacles and draft contingency plans for midcourse adjustments. Hence, the capacity to confront, assess, and make those tradeoffs might be built in advance.

Effective advance planning for such contingencies requires a systematic method. Richard Elmore's 'backwards mapping' approach can be useful: listing all the elements one would need to be working together once an operational system has been assembled, and then planning how to acquire them (Elmore 1979). One of us (Bardach and Patashnik 2016) recommends postulating certain failures (e.g., huge delays, complete program collapse, and bureaucratic resistance) and then writing, from some vantage point in the future, scenarios about how they occurred.

Here, we suggest simulation modeling as a useful implementation planning tool. We use the system dynamics (SD) approach because it is more than merely a quantitative tool for generating internally consistent projections. It is a method of scientific inquiry that helps develop an intuitive grasp of the functioning of complex systems. Compared to less formal approaches, it can help planners anticipate both intended and unintended effects of policy options. First, modeling insists on confronting implementation details often overlooked by policy designers. Secondly, many important details become visible only when the implementation of the policy at hand intersects with other systems within the larger governance context, e.g., procurement rules that severely limit management options or cause delay, local zoning ordinances that obstruct construction plans, and expenditure rules that preclude advancing payments to contractors before work is performed. Formal modeling forces designers to try to analyze what is admittedly a very uncertain field of forces. Thirdly, when bureaucracies become involved, it is often hard to know what will be happening within their sometimes opaque and unpredictable worlds. Certain general outlines can be theorized, but a lot depends on the details of personalities in government positions. Again, the modeling exercise insists on making explicit guesses about the relevant bureaucratic behavior. Finally, systematic modeling makes various value dimensions more visible than they might

²A light illustration of responsibility avoidance is Will Rogers' facetious suggestion during World War I that the best way to fight enemy submarines was to boil the Atlantic Ocean. When asked how that might be done, he replied, "I'm a policy man. I let others worry about implementation" (cited in Wheat 2010).

otherwise be. At the design and adoption phase, one naturally thinks about costs and effectiveness. But as one moves toward implementation, one has to think about delays, goal erosion, and rent-seeking. Formal modeling does not guarantee that unpleasant surprises can be avoided, but it enables policy designers to use a model as a training ground—practice implementation, experience setbacks, and test redesign strategies—in ways that might later prove useful to street level implementers; e.g., see Wheat (2015).

Implementation analysis begins with a definition of the policy to be implemented. At a minimum, the definition should include (1) the nature of the policy mandate intended to accomplish something through the use of a government program, (2) an agency that will take the lead in the activity, and (3) some resources accessible to the agency. Typically, the lead agency will have to *assemble* program elements from other agencies, both public and private, into an operating system—the intended output of the system being, for example, a stream of subsidies or compliant behaviors.³ This assembly process has three main streams. One is *technical*: the elements that need to be put in place to operate a program, such as personnel, organization, office space, manuals, training, clients, hardware, and procedures. Exactly what these elements are will depend on the particulars of the program. The second stream is *administrative*: authority to hire personnel, to expend budget dollars, to procure equipment, and so on. This stream supports activities in the technical stream, though perhaps with some friction and delay, because it proceeds somewhat independently, by its own logic and according to government-wide rules designed in large measure to prevent waste, fraud, and abuse. Thirdly, the *political* stream contains the support or approvals, in their great variety of forms, needed or useful for legitimating a government activity even after a general approval has been given for a policy or project. Given the US federal form of government, this often means that federal agencies seek general cooperation or acquiescence from their state and local counterparts (and constituencies they represent) and, in some cases, from private-sector partners.

10.2 Qualitative Implementation Modeling

The prospect of dealing with mathematical equations is not appealing to many who are engaged in the policy design process. This can cause resistance to using formal simulation models during that process. One way to lower that barrier is to begin with models that are qualitative rather than quantitative. A diagram of an SD model is a conceptual map that can be explored by policy designers without the cognitive burden of mathematics. Such a diagram is a qualitative model of a social or economic structure, including proposed structural changes, i.e., including policy options. It can be used for preliminary feasibility testing of policy proposals by

³Bardach (1977) develops the concept of an implementation assembly process.

encouraging analysts to envision policy outcomes—intended and unintended—and question how a policy would work in practice. When problematic feasibility issues are identified, planners can discuss ways to redesign the policy to improve the feasibility and raise the chances for successful implementation. The result is a revised conceptual model that reflects rejection or revision of initial options, hopefully with justifiable expectations of a more feasible plan for addressing the policy issues.

Qualitative feasibility testing begins by studying a diagram of a proposed policy and raising questions about it. The intent is to brainstorm political, administrative, and technical constraints that might impede the policy's adoption or prevent a policy from achieving its desired outcomes without negative side effects and then suggest ways to redesign the policy to improve its feasibility. This has proven to be an effective way to sensitize future policy designers trained in SD modeling. Students at the University of Bergen use this method in a master's level policy design and implementation course, while learning how to build implementation structure into their models and how to conduct feasibility analysis alongside cost-benefit analysis (Wheat 2013). Figure 10.1 displays a diagram that will be used to illustrate qualitative feasibility testing. The policy issue concerns regulation of over-fishing in a coastal region, and the model is adapted from Morecroft (2007).

The small inset diagram in Fig. 10.1 shows the historical downward trend in the fish stock, plus two alternative futures: continued decline or stability at a higher level. Symbols in the diagram illustrate the three building blocks of SD models: stocks, flows, and feedback loops. The boxes represent stocks (ships at sea and in the harbor, plus the fish population). Flow icons are the 'pipelines with valves' that control the rate at which material moves in and out of the stocks. Feedback effects are illustrated by arrows that form closed circles of mutual causation.

In this example, the policy feedback loop would regulate the number of ships at sea to achieve the desired fish stock. Government regulators would set a target for the number of ships at sea, based on estimates of the fish stock and a comparison with the desired stock. When the fish population is threatened by 'too many' ships at sea, some would remain docked in the harbor. When the situation improves, ship owners would be permitted to take more ships to sea.

After studying the model diagram, the policy design task is to identify political, administrative, and technical feasibility issues that might occur if such a policy were proposed or adopted. Below is a sample of the kinds of feasibility questions that inevitably arise during implementation analysis of the qualitative policy model displayed in Fig. 10.1.

Political Feasibility Issues

1. Does the public generally accept this kind of government regulation of business activities?
2. Will ship owners obey the regulations? Will they interfere with enforcement?
3. Will the government pay for ships sitting in the harbor?

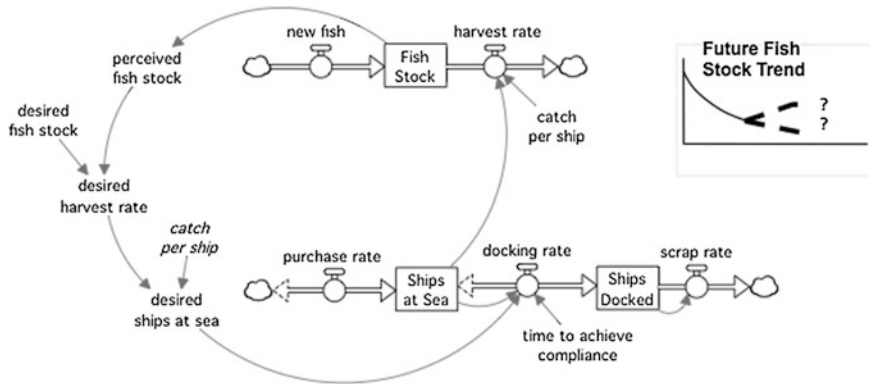


Fig. 10.1 Qualitative policy model of fishing regulation (simplified adaptation from Morecroft 2007, p. 347)

4. What groups are likely to oppose this policy?
5. If idle ships mean idle fishermen, how does that affect the local economy? Will there be pressure for government compensation?
6. What about the ships sailing under a foreign flag? What is the geographic boundary for this policy? Will this policy conflict with existing treaties or trade agreements?

Administrative and Technical Feasibility Issues

1. Which agencies are responsible for estimating (perceiving) ships at sea and the harvest rate? How reliable are their estimates, and what kind of delays should be expected?
2. Which agencies are responsible for estimating (perceiving) the fish stock? How reliable are their estimates, and what kind of delays should be expected?
3. Who will decide desired fish stock? Will the decision be based on an accepted scientific theory? Is there a 'scientific consensus' on the answer to this question?
4. Which agencies are responsible for deciding which ships remain in the harbor? How are those decisions made?
5. Do the agencies have adequate resources (funds, personnel, technology, experience) to do their various tasks?

Brainstorming feasibility questions in the context of a specific policy design is a sensitizing activity. It raises awareness of the potential for policy resistance during both the adoption and implementation stages, and it emphasizes that 'in a system, you can't do just one thing.' The *designed output* of the policy might be a precisely worded set of regulations aimed at a *single desired outcome*. Yet, the exercise

reveals the potential for *multiple actual outcomes*, some of which could lower political support during the policy adoption stage or undermine achievement during implementation.

Qualitative implementation modeling may be sufficient to enable planners to redesign policy proposals in order to reduce chances for disappointing outcomes, or to narrow the number of promising policy options to a feasible subset. For some complex issues, however, simulation modeling can add value to the qualitative approach by quantifying cause-and-effect relationships implicit in a policy idea and projecting the likely behavior that would emerge over time. Moreover, the range of policy outcomes may be particularly sensitive to uncertain assumptions in the minds of policy designers, and simulation modeling enables testing the behavior of a model under various assumptions.

In short, while both types of models can represent the *structure* of a policy, only a quantitative simulation model permits analysis of the dynamic *behavior* that is expected to arise from that structure. In the next section, we demonstrate how quantitative modeling can aid the policy planner, and we use the Oakland case to provide a real-world context for a stylized simulation model of policies aimed at local economic development.

10.3 Quantitative Implementation Modeling

We approach the building of the simulation model from three directions. First, we rely on available empirical evidence which, in this case, consists of a well-documented case study of an implementation process to help ground our model in at least one actual instance. This provides structural and behavioral benchmarks against which to compare our model's structure and behavior. Here, our benchmarks are provided by Pressman and Wildavsky's case study of a US federal policy initiative to increase hiring of long-term unemployed persons in Oakland, California, during the 1960s.⁴ Secondly, we rely on our general theoretical understanding of governance and political processes. For example, we assume that government agencies typically specialize by mission—turning out grants to businesses, for instance, or guarding the integrity of procurement decisions—and tend to emphasize the priority of that mission at the expense of other values that, when balanced properly against the mission priorities, might deserve higher weights than they receive. Finally, we conceptualize as stocks and flows the variables suggested by our theoretical and empirical foundation, and define the boundary of the model broadly enough to reveal an endogenous feedback structure that accounts for the behavior of the model.

⁴The Pressman and Wildavsky book is the sole source of facts about the Oakland case, although their case study has generated analyses too numerous to count (e.g., a Google search for "Pressman and Wildavsky" yields 15,000 hits).

10.3.1 *The Oakland Story*

Policy designers are habitual optimists. The world where the policy will be implemented is, by nature, less hospitable to the designers' wishes than they would like to believe. Things cost more, take longer, and are more subject to being hijacked by political interests who do not care much about the original policy objectives but do care a lot about their own policy, institutional, and career interests. With only occasional exceptions, therefore, the implementation phase of policy-making is disappointing. And the story of EDA in Oakland is not one of those exceptions.

The Oakland case is an old one.⁵ Yet, it suits our purpose for two reasons. First, it is well known for its illustration of implementation issues that are uniquely problematic in the public sector, namely those requiring reconciliation of diverse public and private interests and coordination of multiple bureaucratic programs and procedures. Another reason is its special relevance to an outcomes shortfall: it was a jobs-for-hardcore-unemployed project that cost more than \$10 million but created fewer than 100 jobs, far from the goal of 3000. Moreover, few if any of the jobs went to the target population. Another \$13 million was scheduled for spending, but the plug was pulled on the Oakland project before the wasted effort could escalate even higher.

The seeds of the project had been planted in 1965, when the US Congress authorized and funded a government subsidy program for public works projects that would support local economic development designed to encourage hiring long-term unemployed persons, most of whom were racial minorities. The lead agency was the Economic Development Administration (EDA) in the US Department of Commerce, and EDA focused its resources on Oakland, California. A local public agency, the Port of Oakland (the Port), would receive the federal government funds and build an airplane maintenance hangar, which it would lease to World Airways (World). In effect, EDA was contracting with World through a public-sector intermediary. In return, World was expected to hire local unemployed persons for the short-term construction jobs and for the more skilled long-term maintenance jobs. The expectation was that EDA and World would jointly arrange for the training of job seekers and new hires. The needed technical elements to be assembled in Oakland were: (1) jobs, (2) qualified potential employees; (3) a way for government to enforce hiring commitments by recipients of the funds; and (4) training for a large fraction of the potential employees.

⁵As are the authors. One of us was literally present at the creation of the Oakland case study project led by Pressman and Wildavsky at Berkeley, having been a professor of public policy at the Goldman School of Public Policy since 1970. At that time, the other author was a student of public policy at Harvard's Kennedy School, thereafter serving on the White House staff. We have seen our share of gaps between policy efforts and outcomes, not only in academic research but also while in government staff positions and as consultants to governments.

Despite the availability of EDA funds amounting to \$23 million in 1966, signs pointing to a disappointing outcome were evident early in the project, as various delays ensued. During lengthy contract negotiations with EDA, World objected to any provision that would permit EDA to reclaim funds contingent on post hoc approval of World's hiring successes; in the end, World would agree only to including a *plan* for hiring in the initial contract. In 1968, the Port estimated a cost overrun of nearly \$5 million for the hangar project and asked the EDA to absorb it. EDA tried to use the occasion to leverage its demands on World to further the hiring and affirmative action goals, and continued to do so through early 1969, when it finally turned down the Port's request. Meanwhile construction did not go forward. On at least one occasion, World apparently threatened to back out of the project if the EDA put World at greater financial risk. Early in 1969, World told EDA that it was withdrawing its hiring plan in favor of one that promised less minority employment.

The worker training program never materialized. The program needed numerous approvals: by World, by units within the US Department of Labor and the US Department of Health, Education, and Welfare (HEW), by the California state Department of Employment Development, and by EDA. Reviews and negotiations went on for nearly 2 years, until HEW finally vetoed the plan in 1968 and World ceased participation in plans for worker training.

The contracts for architectural plans for the hangar were not let until mid-1971, nearly 6 years after the initial mandate, and fully 5 years after the EDA had made a big public announcement that it had a project on track that would produce 3000 jobs in the Oakland labor market. In the end, the number of new jobs totaled only 2–3% of that goal, and only a small fraction of that total went to the target group: long-term unemployment persons.

10.3.2 Behavior of the Model

The SD modeling process usually begins by studying a time series graph that displays historical patterns of behavior that a model will be designed to explain. However, despite several careful readings of *Implementation*, all we can say for sure is that the \$23 million of EDA funds were not fully distributed during the 6 years from 1966 to 1972, a period within which most observers expected the investments to be made. Cumulative spending was closer to \$10 million. And the number of *new* jobs created was nowhere near 3000; in round numbers, it was probably no more than 100, if that many. We want to compare these rough historical estimates at the end of 1971 with the simulated results generated by our model.

Comparing model behavior with even rough estimates of historical Oakland 'data' requires calibrating our generic model with numerical estimates or, in some cases, guesstimates of Oakland-relevant parameters. Given what we know about the Oakland case, we can safely assume that training capacity did not exist and that the

total number of jobs actually created in the Oakland project was no more than 100. Therefore, in the model, we set training capacity equal to zero and the initial value of qualified long-term unemployed persons (those not needing to be trained) at 100. Some of the other parameters in the model are not necessarily Oakland-specific, although we attempted to base estimates on empirical economic data for Oakland during the 1960s whenever possible.

The results are displayed in Fig. 10.2, which compares the simulation results with our knowledge of cumulative spending and employment. The thin lines indicate simulation results and the wide bars represent the data estimates for cumulative EDA spending and employment.⁶ We made no attempt to speculate about the unknown historical pattern; thus, the bars show the best guess total at the end of the project.

The top frame indicates that the simulated cumulative spending after 6 years is similar to the 'data' we have (about \$10 million). Likewise, the bottom frame shows a simulation result that is consistent with the upper bound estimate of new jobs (100) actually created by the Oakland project.

The simulation experiment described above, while pertinent to the circumstances in Oakland, does not permit exploring the full range of behavior our model can generate, primarily because we assumed zero training capacity. We will now reverse that assumption and observe how strategic interaction between government agencies and private-sector institutions can generate a range of plausible behaviors when training capacity is optimal. The interaction in the model can be aggregated and summarized as the degree of company *cooperation* with the government. In this context, full cooperation includes a shared goal for total project employment and the time period during which that goal should be achieved. That would mean, for example, company acceptance of a target capital-labor ratio that would be lower than the company's normal target. In our model, that has implications for a company's willingness to adopt the government's 5-year employment goal and the short-term employment targets; and the latter has immediate impacts on hiring. The desired pace of company investment may also conflict with the government's deadlines. These sources of conflict do not necessarily have to be activated; they can remain dormant and, if they do, we will call that 'company cooperation' with the government. Conversely, a lack of alignment between the goals of the company and the government constitutes lack of cooperation.

The company's response to government sanctions is also indicative of the degree of cooperation. If the company falls behind the government's desired hiring rate and pays a penalty in terms of slower cost reimbursement, cooperation means that the company acknowledges its failure and the legitimacy of the penalty and does not

⁶In the model, *LTU Employed* refers to long-term unemployed persons actually hired, and that is the variable graphed in Fig. 2. However, we should emphasize that whatever the actual employment total in Oakland, only a fraction of that number included the target population, and this discrepancy is not specified in our model. In addition to assuming no training capacity, the simulation results in Fig. 2 also assume weak cooperation between World and EDA, the interpretation of which is explained in the text.

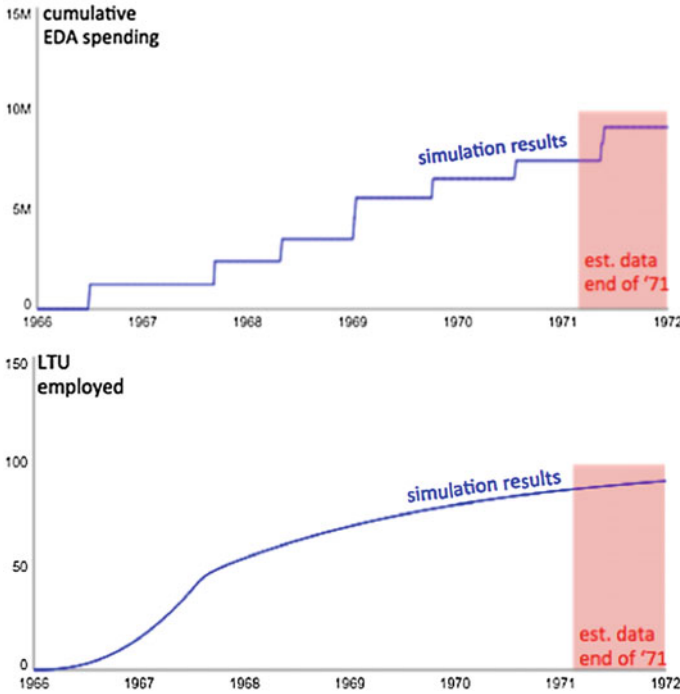


Fig. 10.2 Model behavior and estimated Oakland data (historical pattern of data unknown)

retaliate in any way. In the model, retaliation by an uncooperative company takes the form of slowing the hiring rate. Figure 10.3 displays the results of three simulation runs, each with different assumptions about company cooperation. Note that the simulation continues beyond the 5-year government subsidy program; thus, this should be viewed as a generic test of model behavior that has nothing to do with the details of the Oakland case even though the horizontal axis still refers to that time period in history.

As before, *LTU employed* refers to total project employment. The *Target LTU employment* refers to the company's goal, which matches the government's goal only when there is full cooperation. The best-case scenario (top frame) requires optimal training conditions (capacity to train 500 persons per year, at least 20% enrollment potential each year, 100% training success, and no dropouts) plus full company cooperation. That scenario generates employment that approaches the government's goal, but it takes more than a decade to do so, despite tacit company acceptance of the government's hiring schedule. Failure to keep pace with that schedule results in government sanctions (delays in cost reimbursement), but the full cooperation assumption assures no retaliation in this scenario and, eventually, the government's desired employment level is reached. With weak or nonexistent company cooperation (middle and bottom frames), employment stabilizes below

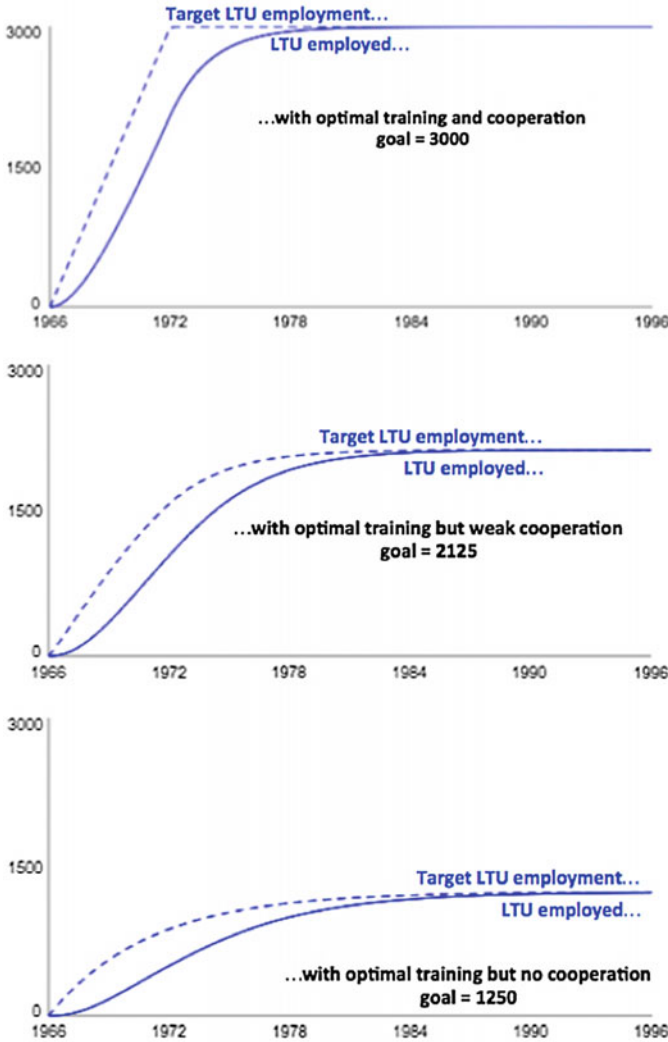


Fig. 10.3 Growth toward rising goals

government’s goal even if there is optimal training capacity. Despite the quantitative differences in Fig. 10.3, there is a similar qualitative behavior in all three frames: goal-seeking patterns for both the target and actual employment levels. Employment rises toward a rising employment target. Next, we examine the structure of model, seeking the source of these persistent dynamic behavior patterns.

10.3.3 Structure of the Model

The full model consists of four sectors: hiring, training, spending, and reimbursing. Before examining the detailed stock-flow-feedback structure in each of those sectors, we present a high level view of the feedback structure responsible for the goal-seeking behavior pattern displayed in Fig. 10.3. A simple set of feedback loops ties together three sectors of the model: hiring, company spending, and government reimbursement. The feedback loop diagram in Fig. 10.4 displays the source of the goal-seeking dynamics in the model.

Feedback loops are distinguished by their positive or negative polarity. Positive feedback loops have self-reinforcing effects. There is no normative connotation in the ‘positive’ label; behavior that feeds on itself can cause growth or collapse and, depending on one’s values, can be virtuous or vicious. To avoid a misunderstanding, positive loops are often called *reinforcing* loops, denoted in feedback loop diagrams by the letter R. In contrast, negative feedback loops have self-adjusting effects. Their goal-seeking structure counteracts tendencies for a system to grow or collapse. Sometimes called *counteracting* loops, they are denoted by the letter C.

The feedback loop diagram in Fig. 10.4 reveals two counteracting loops, C1 and C2, that are responsible for the goal-seeking behavior in the full model, and a reinforcing loop R1 that has the potential to weaken loop C1 and hinder its goal-seeking tendency.

The hiring loop C1 functions in a way that closes any gap between target and actual employment. The faster the hiring adjustment time, the quicker the gap is closed. Previously, we discussed the potential for government to seek leverage over the company’s hiring process by slowing the reimbursement process. When actual employment fails to keep up with the government’s scheduled employment goal, the reimbursement time increases. The company’s retaliation option is to slow the hiring process even further. That is the essence of loop R1 when activated by sanctions and retaliation; it can frustrate both the government and the company and, in so doing, weaken the net hiring loop C1.

Target employment depends on the company’s stock of physical capital (infrastructure, equipment, tools, etc.) and the desired capital-labor ratio. If investment exceeds depreciation (not shown), the company’s capital increases and the target for employment increases proportionately. Growth in the capital stock is controlled by loop C2, which closes any gap between actual and target capital. To the extent that the company aligns its operating strategy with government’s policy goals, the desired capital-labor ratio, the target for capital, and the pace of adjustment—and, therefore, target employment—would reflect the government’s goals. Lack of company cooperation would reduce alignment with government’s goals, lower the target employment, and reduce the hiring rate in loop C1. These strategic interactions between government and the company are exogenous in the current version of our model. The degree of goal alignment can be varied by the user of the model and the impact of different assumptions can be observed in the simulation results.

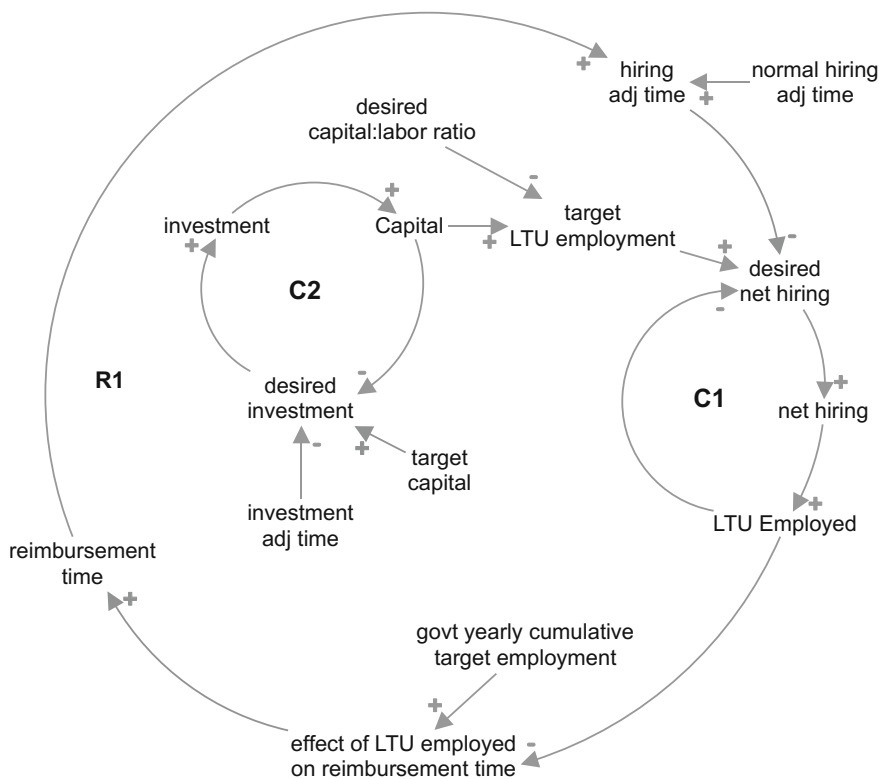


Fig. 10.4 Feedback loops responsible for goal-seeking behavior extracted from full model in Fig. 10.10

Our hypothesis that loops C1 and C2 are responsible for the goal-seeking behavior is supported by two simulation experiments with the full model.⁷ Figure 10.5 displays the model’s behavior when loops C1 and C2 were deactivated or ‘cut’ during the simulation. In the left frame, cutting loop C2 stops investment and the growth of the capital stock which, in turn, stops the growth in *Target LTU employment*. In the frame on the right, cutting C1 stops the growth of *LTU Employed*. The employment target is not part of that loop and continues to rise to its own goal, unaffected by the deactivation of loop C1.

In the remainder of this section, we examine the details of the model’s stock-and-flow structure and gain additional insight regarding the source of dynamics in the model. Figures 10.6, 10.7, 10.8, and 10.9 display close-up views of the four sectors in the model, and the full model is displayed in Fig. 10.10. Although the ‘EDA in Oakland’ case motivated the model, we have adopted

⁷For this test, a training program is activated so that the stock of qualified applicants is large enough to accommodate the desired hiring rate.

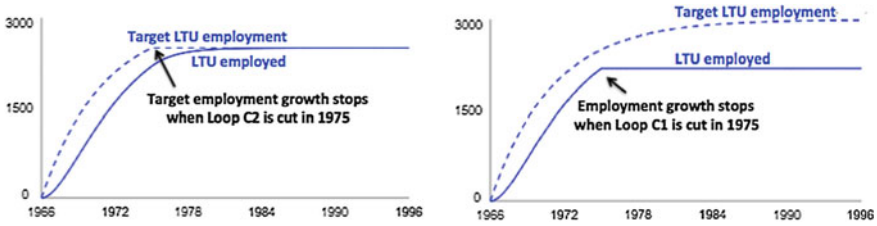


Fig. 10.5 Cutting counteracting feedback loops stops goal-seeking growth

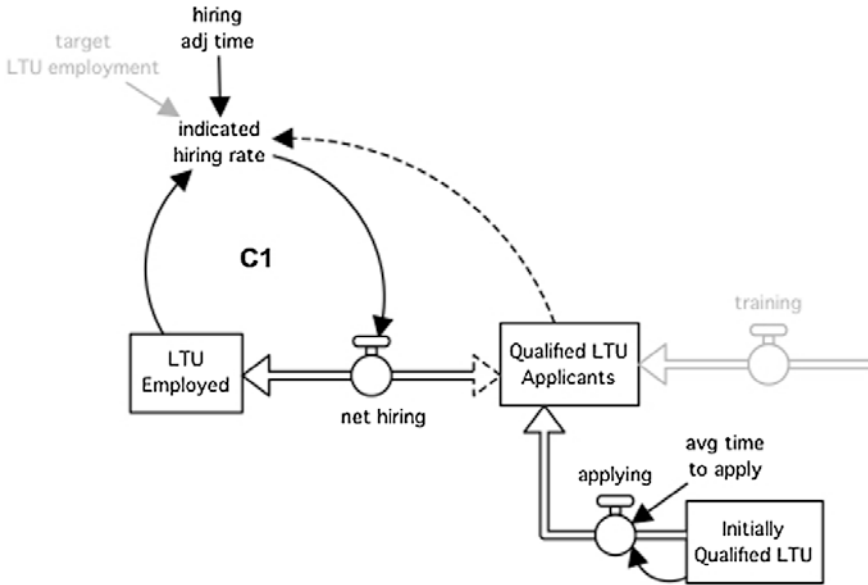


Fig. 10.6 Hiring sector

generic names such as ‘government’ for EDA and ‘company’ for the various private interests, the largest of which was World Airways. We have also selected round numbers for parameter values such as delay times and various coefficients in the model affecting spending, training, etc. All parameter values can be modified by users wanting to test the effects of different assumptions. The generic approach facilitates adapting the model for other policy design research tasks, and using it as a ‘method of inquiry’ tool for policy designers.

Figure 10.6 displays the stock-flow-feedback process that governs hiring in the model. As long as *target LTU employment* exceeds *LTU Employed*, *Qualified LTU Applicants* are being hired. When *net hiring* is negative, layoffs occur. The faster the *hiring adjustment time*, the sooner actual employment rises to meet the target. This is the same counteracting loop C1 displayed in Fig. 10.4. Here, however, the stock-and-flow structure specifies how the process operates; what Richmond (1994)

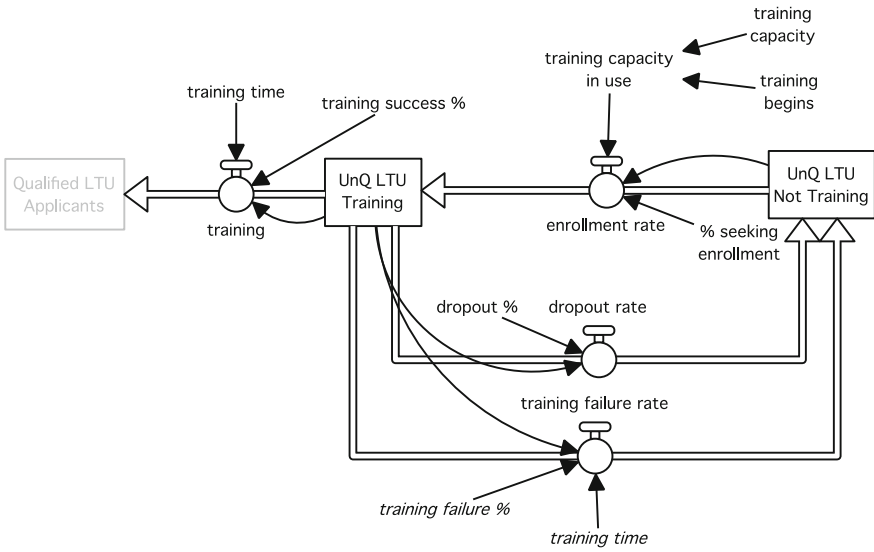


Fig. 10.7 Training sector

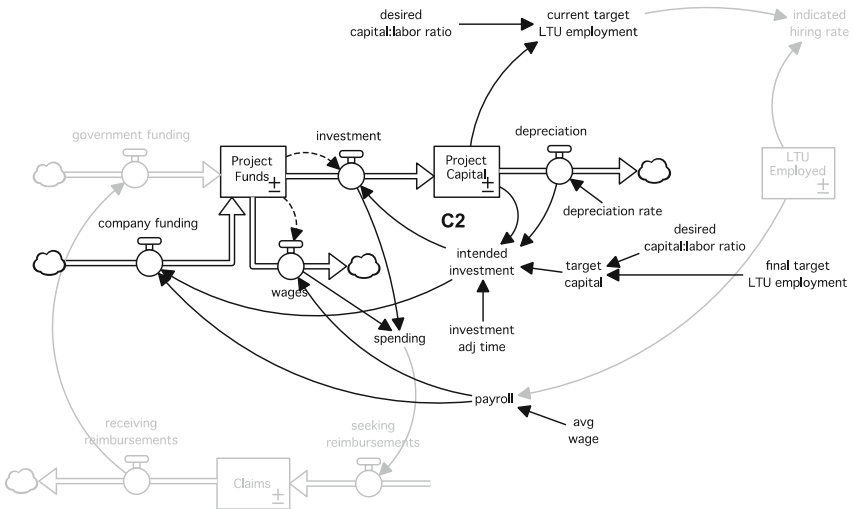


Fig. 10.8 Spending sector

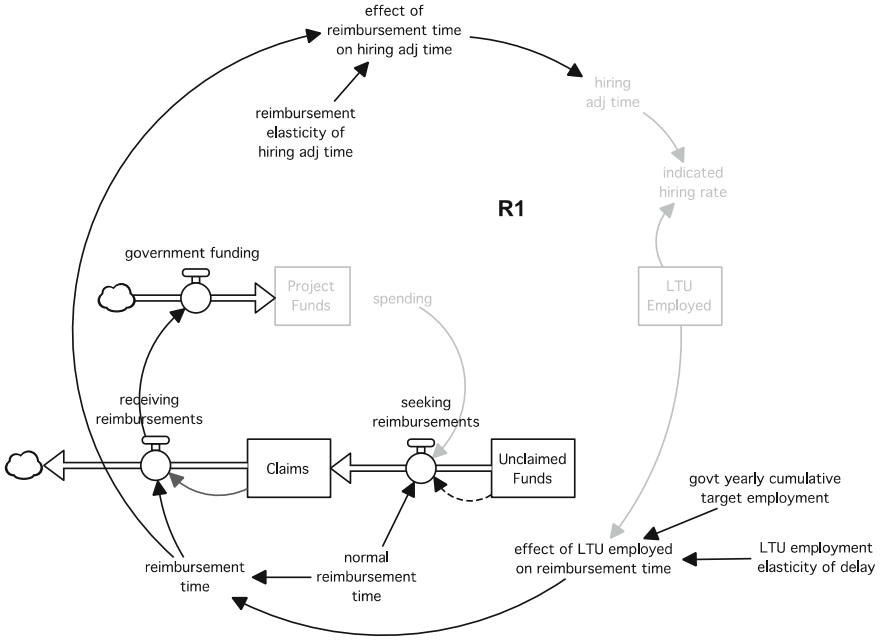


Fig. 10.9 Reimbursing sector

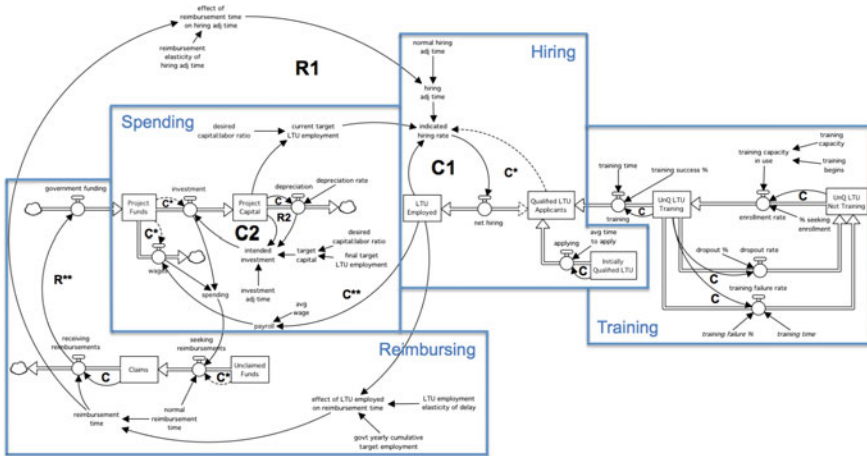


Fig. 10.10 Simplified view of full model

calls the ‘physics’ or ‘plumbing’ of the system. Significantly, it reveals the real-world constraints on hiring. Hiring requires a stock of *Qualified LTU Applicants* (initially zero) that depends on an inflow of applicants from the *Initially Qualified LTU* (assumed to be 100) or those successfully completing their *training*. As long as there are qualified applicants, loop C1 operates freely. Otherwise, the dashed link signals the absence of qualified applicants and the *indicated hiring rate* is zero, making loop C1 dormant.

Figure 10.7 displays the training sector of the model and reveals its connection to the hiring sector, via the *training* flow. Although training never materialized in Oakland, this sector is an essential component of any model of a job-creation policy because it raises critical policy design questions. The annual *training* rate depends on the number enrolled in a training program (initially zero), the time it takes to train them, and the fraction successfully trained; i.e., those truly qualified and available for employment. Those failing to be trained rejoin the ranks of the unqualified LTU not enrolled in a training program (6000 initially, based on rough estimates for Oakland in 1965). In addition, there are dropouts. The annual *enrollment rate* depends on the capacity of the training facilities and the percentage of LTUs enrolling each year. In the Oakland story, *training capacity in use* is zero, which prevents enrollment and training and (in Fig. 10.6) hiring. In other cases, training capacity may exist but insufficient enrollment, high dropout rates, or ineffective *training* may limit growth in the number of qualified applicants. Each of these leverage points should be highlighted during the policy design stage to activate contingency planning.

The spending sector is displayed in Fig. 10.8, along with its connections to the (dimmed) hiring and reimbursement sectors. Company *spending* is the sum of *investment* and *wages*, and the total drives reimbursement *Claims*. To jump-start the process, *company funding* is needed, but *government funding* replenishes the *Project Funds* stock as reimbursements are received. The dashed links to *investment* and *wages* slow those outflows if funds run low, and no *spending* occurs if there are no funds at all. *Investment* adds to *Project Capital*, in response to feedback loop C2 that gradually adjusts the current capital stock to its target value.

Both the capital target and adjustment time are influenced by strategic interaction between the company and the government. With full cooperation from the company, the *desired capital-labor ratio* and therefore, *target capital*, will reflect the *government’s final target LTU employment*. With company resistance, the target will more likely resemble the company’s capital-labor ratio preference. Likewise, the degree of company alignment with the government’s project deadline determines the time period over which the capital stock is adjusted (in the model, the particular strategic reactions are exogenously controlled by the user, and the controls are not shown in Fig. 10.8). For private companies, demand for labor is usually derived demand; i.e., it depends on the demand for the goods and services that labor can produce. Here, we simplify the labor demand structure by assuming the company regularly adjusts its target for employment based on the level of installed capital and the (exogenously determined) *desired capital-labor ratio*.

The *current target for LTU employment* then influences hiring, and changes in *LTU Employed* affect the *payroll* and the next round of *spending*.

The final part of the model to inspect is the reimbursing sector, displayed in Fig. 10.9. This sector governs the reimbursement process after the company submits a project spending claim. This sector interacts with both the spending and the hiring sectors (both partially displayed and dimmed). In the Oakland project, EDA distributed funds only to reimburse company spending after the fact. One could imagine other possibilities, but that is not an uncommon way that governments distribute grants; thus, it is the procedure we assume here. We also assume the government slows the reimbursement process during periods of negotiation when the company fails to meet government's annual hiring targets (estimated as a linear trend from the beginning to the end of the project). As discussed previously, feedback loop R1 implements the company's retaliation when reimbursements are late. The effect of the loop is to lengthen the hiring adjustment time, further slow the employment of LTUs, and reinforce a vicious mutual effect on the government, the company, and the long-term unemployed persons waiting to be hired.⁸

Figure 10.10 displays a simplified version of the full model, with several parameters and one flow (company funding) deleted for clarity. Close scrutiny reveals 16 feedback loops, 13 of which are counteracting, and only those could account for the goal-seeking behavior generated by this model. The four denoted as C* (with a dashed link in the loop) are dormant unless their relevant stocks approach zero.⁹ Six of the remaining counteracting loops have an implicit purpose of draining their stocks to zero; none could be pushing employment up toward a goal. For example, the training sector's counteracting loops constrain hiring; the cumulative net inflow to *Qualified LTU Applicants* represents the maximum number that could be hired but that number does not drive the hiring rate. That leaves only loops C1 and C2 as the source of goal-seeking dynamics, with loop R1 weakening the employment adjustment impact of loop C1, as confirmed by our previous analysis (Figs. 10.4 and 10.5).

⁸The strength of loop R1, assuming it is activated, depends on assumptions about the reaction functions influencing the government and the company. For example, we assume the government increases the normal reimbursement time by 3% when *LTU Employed* is 10% below the government's target level (elasticity = -0.3). We assume the company slows the hiring adjustment to match the slowdown in the reimbursement process (elasticity = 1.0).

⁹The reimbursement loop R** aggregates two loops—one stemming from *wages* and the other from *investment*. However, R** never becomes a closed loop unless the C* loops are active, in which case *Projects Funds* would be zero. If R** raised *Project Funds* above zero, that would make the C* loops dormant and immediately deactivate R**. The *Project Funds* stock constrains spending on *investment* and *wages* but it does not drive those outflows. Similarly, the potential C** payroll loop has no effective feedback effect on *LTU Employed* because the loop is only closed when *Project Funds* is at or near zero. We include R** and C** in our total feedback loop count, but they could not be responsible for the model's goal-seeking behavior.

10.4 Conclusion

Qualitative modeling can sensitize policy designers to the technical, administrative, and political feasibility issues that can impair policy initiatives with time-delayed destructive elements. Quantitative simulation modeling can add value to qualitative maps by revealing the dynamics of complex systems, and experimenting with a simulation model provides vicarious experience in policy design and can hone the skills of policy designers.

The questions raised by the qualitative fishing regulation example (Fig. 10.1) illustrate how implementation difficulties can be predestined by the original policy design. And, in the Oakland example, the diagrams in Figs. 10.6, 10.7, 10.8, 10.9 and 10.10 could be used to generate questions about make-or-break issues such as the training program or to anticipate the likelihood and implications of divergent company and government goals or the company's reaction to government sanctions and the likely impact of that vicious circle on the pace of employment. A collaborative effort to sketch a causal model of how a policy is expected to work is likely to generate critical questions about policy ideas. A policy design tool that provokes this kind of thinking and communication promises to be useful to those with responsibility for envisioning outcomes.

Quantitative simulation models encourage planners to view feasibility issues in the context of activity streams that flow over time, interact in unexpected ways, and generate outcomes that may not be intended. The Oakland model, for example, demonstrates how millions of dollars could be spent before it becomes apparent that no training program would materialize. Witnessing a stream of spending that does not produce jobs could energize efforts to make sure that obstacles in the way of training would receive early and continuous attention. Simulation experiments also reveal (in Fig. 10.3) that a training program is a necessary but not sufficient component of a jobs-creation project. Without company cooperation, the employment potential could be well below the government goal even with optimal training capacity. Moreover, formal methods of quantitative model analysis can identify the structural reasons for dynamic behavior (Figs. 10.4 and 10.5) and provide valuable clues about how to redesign a process to achieve a better outcome. For example, simulation results reveal how the company retaliation feedback effect (loop R1) can undermine the hiring process (loop C1). If apparent during the policy design stage, such results could foster debate about the potential for certain types of sanctions to be counterproductive, and a model could enable tests of alternative ways to sanction. Even without further testing, the simulation results could raise the debate about sanctions to a higher level of specificity about how they would work, the reactions they might provoke, and the expected impact on outcomes. Simulation results in our example also underscore the critical importance of alignment between company and government goals regarding employment targets and desired levels and timing of investment, and reveal the naiveté of simply assuming that subsidies would result in company operations that followed government guidelines instead of standard business guidelines and procedures.

We encourage policy designers to look for synergy in the joint use of these approaches with other good methods; for example, the failure scenario writing exercise described in Bardach and Patashnik (2016). Designed to brainstorm ideas for disaster avoidance or damage control, that exercise can assess the feasibility of a policy option. When used in combination with qualitative feasibility testing, it would encourage mental simulation of unintended consequences. In addition, qualitative feasibility testing specifies implicit mechanisms in a policy, and that can enrich the scenario writing process by spotlighting the specific resources that must be assembled to facilitate implementation. The value is not in a model per se; the value is in how the modeling process can shape the mental models of the participants in advance and thereby influence their strategic thinking, their contingency planning, and their design of the content and transmission mechanism of a particular policy.

We acknowledge limits to implementation modeling. Modeling is no panacea for policy failures in public institutional settings characterized by conflicting views and shared powers. We do not think that everything about a policy that *might* be modeled *should* be modeled. Certainly, not all implementation-relevant factors are included in the model inspired by the Oakland story. Some of the limits are deliberate. Like a highway map that omits local streets, the details of a simulation model reflect its purpose, and a high-level model of a job-creation program will permit later addition of contextual details. Other limits are problematic. For instance, when considering how to model discrete as well as continuous patterns of political conflict among officials who share powers within and across governmental units, an argument can be made for an agent-based approach. Yet the more aggregated system dynamics approach is better for mapping endogenous feedback structure and encouraging operational thinking about how complex systems work and how they could be modified to work better. In this example, a methodological compromise may be justified and is certainly possible.¹⁰

We envision an accessible inventory of generic but insightful causal models that can be adapted for practitioners in the policy design arena. Developing such models requires closer collaboration than currently exists between the modeling disciplines and the public policy research disciplines, something we have encouraged (Wheat 2010; Wheat and Bardach 2015). A desirable by-product of such collaboration would be a new instrument in the research toolkit that policy analysts could use to improve understanding of gaps between policy inputs, outputs, and outcomes.¹¹

¹⁰For example, *AnyLogic* (anylogic.com) software supports both agent-based and system dynamics modeling. Moreover, one of our colleagues at the University of Bergen, Pål Davidsen, is using features of *Stella Architect* (iseesystems.com) to represent individual agents interacting within a system dynamics model.

¹¹The Oakland model is available for online simulation at <https://sims.iseesystems.com/david-wheat/oakland/#page1>. Readers wishing to use *Stella Architect* to study model equations and experiment with alternative formulations are encouraged to request a fully editable copy of the model from the authors.

References

- Bardach, E. (1977). *The implementation game: What happens after a bill becomes a law*. Cambridge: MIT Press.
- Bardach, E., & Patashnik, E. (2016). *A practical guide for policy analysis*. (5 th ed.). University of California, Berkeley Washington: CQ.
- Elmore, R. F. (1979). Backward mapping: Implementation research and policy decisions. *Political Science Quarterly*, 94(4), 601–616.
- Morecroft, J. D. (2007). *Strategic modeling and business dynamics: A feedback systems approach*. Chichester: Wiley.
- Pressman, J. & Wildavsky, A. (1973). *Implementation: How great expectations in Washington are dashed in Oakland, or why it's amazing that federal programs work at all; this being a saga of the economic development administration as told by two sympathetic observers who seek to build morals on a foundation of ruined hope*. Berkeley: University of California Press.
- Richmond, B. (1994). Systems thinking/system dynamics: Let's just get on with it. *System Dynamics Review*, 10(2–3), 135–157.
- Wheat, I. D. (2010). What can system dynamics learn from the public policy implementation literature? *Systems Research and Behavioral Science*, 27(4), 425–442.
- Wheat, I. D. (2013). *Teaching policy design: A case study of unintended consequences when the EU regulates hospital doctors' hours*. International System Dynamics Conference. Cambridge, USA.
- Wheat, I. D. (2015). Model-based policy design that takes implementation seriously. In E. W. Johnston (Ed.), *Governance in the information era: Theory and practice of policy informatics*. New York: Routledge.
- Wheat, I. D. & Bardach, E. (2015). *Public policy implementation modeling: The classic case of EDA in Oakland*. International System Dynamics Conference. Cambridge, USA.
- Wilcox, R. F. (1978). Review of the implementation game: What happens after a bill becomes a law. *Policy Sciences*, 10(2–3), 225–227.

Part IV
**Applying Outcome-Based Performance
Management: Experiences from Different
Public Sector Domains**

Chapter 11

Evaluation of Innovation Performance in the Public Sector: A Systematic Review of Studies

Greta Nasi, Maria Cucciniello and Virginia Degara

Abstract Innovation promises to produce beneficial results for the most part. Consequently, there has been a great deal of discussion in management literature by academics and practitioners in recent years as to the impacts of innovation, on account of the extensive adoption of innovation in the public sector. The debate is currently open on multiple levels: the return of investment after the adoption of innovation, its contribution to the transformation of public sector organizations, and its impacts on society. The purpose of this chapter, based on a systematic review of literature, is to: discuss the status and trends in measuring innovation impacts; present cases and experiences; and set the basis for drafting a road map for moving forward in this field.

Keywords Literature review · Impacts · Innovation · Performance

11.1 Introduction

Although a large body of literature has focused on the adoption of innovation in the public sector over the last 40 years, only a minor stream of research has investigated the evaluation of innovation performance (Damanpour et al. 2009; Jaeger

G. Nasi (✉)

Department of Policy Analysis and Public Management, Public Management and Policy Department, SDA Bocconi School of Management, Bocconi University, Via Roentgen, 1, 20136 Milan, Italy
e-mail: greta.nasi@unibocconi.it

M. Cucciniello

Department of Policy Analysis and Public Management AND Public Management and Policy Department, SDA Bocconi School of Management Bocconi University, Bocconi University, Via Roentgen, 1, 20136 Milan, Italy
e-mail: maria.cucciniello@unibocconi.it

V. Degara

Bocconi University, Milan, Italy

© Springer International Publishing AG 2018

E. Borbonovi et al. (eds.), *Outcome-Based Performance Management in the Public Sector*, System Dynamics for Performance Management 2, DOI 10.1007/978-3-319-57018-1_11

203

2005; Andrews et al. 2012; Ahn and Bretschneider 2011). However, given the socio-economic contexts in which Public Sector Organizations (PSOs) operate, demanding they deliver more with less resources, more and more attention is being attributed to the need to monitor the impacts of innovation on performance (Cucciniello and Nasi 2013; Walker et al. 2011; Kattel et al. 2013; Bloch 2010; Dunleavy et al. 2009).

Although innovation has also been recognized as a main contributing factor to the modernization of PSOs, measuring the impacts of innovation still represents an open and critical challenge (Kim et al. 2005; Stefanou 2001).

The most recent debate as to the importance of evaluating the effects of innovation came in the wake of President Obama's requests for a smarter, more innovative, and more accountable government for citizens. Like his two immediate predecessors, President Obama has made government management a priority and since taking office, his administration has stressed performance measurement and evidence-based decision-making. However, as the recent history of the implementation of the Affordable Care Act notoriously demonstrated, adopting complex innovation that affects an organization's structure and culture, work processes, behaviour, and communication channels, can be considered one of the most difficult and challenging tasks to overcome in the public sector.

Multiple approaches to the evaluation of innovation have been discussed and presented in literature. Most studies are based on more traditional efficiency and effectiveness models, whereas others suggest expanding their focus and taking more of the social and ideological effects pursued by public sector innovation into account (Moore 1994, 1995). Some authors have also suggested the need to perform stakeholder analysis in order to depict the value of innovation for individuals and understand how this determines and influences the overall impacts of innovation adoption (Dawes et al. 2009).

In the private sector, a substantial body of empirical and theoretically informed research has discussed return of investment measures, and indicators of key performance and success. The main reason for evaluation is the need to monitor profitability results, in turn providing an incentive for further innovation in order to cut costs, improve market share and create new products and services. Public sector innovation shares some parallel goals, such as improved efficiency, productivity and adequacy of programs and services, but the value of innovation in the public sector differs substantially from its value in the private sector and can be more complex and more difficult to measure (Walker 2008).

The lack of focus on performance represents a real obstacle when trying to answer the "so what" question that is frequently put to public managers.

The purpose of this work is to investigate the current status of scientific literature dealing with the measurement of innovation performance in the public sector, devoting particular attention to the evaluation methods applied. The former step is crucial for identifying the difficulties and inconsistencies encountered by the evaluation methods adopted so far. What's more, identifying any current gaps in literature may enable us to create the basis for a comprehensive evaluation method that integrates existing ones.

In this context, this work undertakes a systematic literature review and aims to: (i) discuss the status and trends when measuring the impact of innovation; (ii) present cases and experiences; and (iii) set the basis for drafting a potential road map in order to move ahead in this field.

Our chapter is structured as follows: Sect. 11.2 presents the framework adopted in the study; Sect. 11.3 presents the method used for the literature review and discusses the characteristics of the records resulting from our systematic literature review; Sect. 11.4 analyzes and comments the results emerging from the records and their evaluation methodologies, before suggesting a multidimensional framework for evaluating innovation. Sections 11.5 and 11.6 illustrate and discuss the critical points that emerged as a result of our analysis, and Sect. 11.7 suggests insights as to how to interpret the evaluation of innovation and so move forward.

11.2 Theory

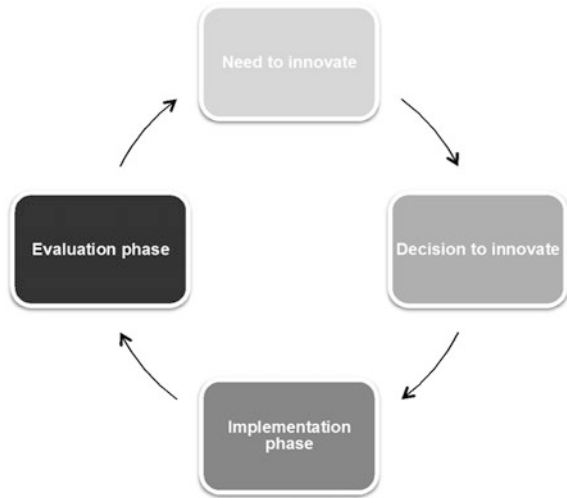
Several studies (such as Van de Ven et al. 2008) found that an innovation process or a set of innovation activities (Damanpour and Schneider 2009) do not resemble a simple linear model. This linear model is very often dominant in more normative and prescriptive innovation models (Bason 2010). However, it has been seen that innovation processes are a rather messy and complex progress of events, pointing in all directions and making use of all sorts of feedback from different stakeholders (Fagerberg 2005) leading to the use of more sophisticated ideas. However, some patterns of similarity in the progress of these events can be observed (Van de Ven et al. 2008, pp. 23–24; also see Rogers 2003; Osborne and Brown 2005; Damanpour and Schneider 2009). According to several studies (Lapsley and Llewellyn 1998), there are many powerful stakeholder groups within health care organizations and each of these can influence the ultimate success or failure of an overall innovation process.

One way to achieve a broader overview of the innovation process could be to look at the innovation life cycle as presented in literature (Mulgan and Albury 2003; Greenhalgh et al. 2004).

This model could build on existing initiatives and extend the overview of innovations to include the main stages of an innovation life cycle (Fig. 11.1).

The first stage of this model focuses on the “*need to innovate*”. For Mulgan and Albury (2003), this means “generating possibilities” or ways to stimulate and support ideas for innovation. Rogers (2003) expresses it as the “knowledge phase” in which the innovation agent becomes aware of the possibility of innovation and the “persuasion phase” when the agent becomes progressively interested in the innovation. In general, the need to innovate emphasizes the importance of generating possibilities. It represents the starting point for innovation: there is an idea that a need is not being met, coupled with an idea of how it could be met. Mapping how this stage occurs might offer policy insights and help stimulate innovation and management actions to enhance innovation in practice. The idea that scientific

Fig. 11.1 The innovation life cycle



knowledge plays a dominant role in this gestation period should be put into perspective. Other sources of innovation seem to be more important, like the needs of customers (Von Hippel 2006).

In this model, the *decision to innovate* represents a second step. According to several studies (Mulgan and Albury 2003), this phase involves taking a promising idea and putting it into practice on a small scale. Understanding how this phase occurs could be useful for policy makers so that they can activate sources of innovation that meet the needs of public service organizations, the expectations of the context, and are compatible with the environment in which they are introduced. It can also contribute to management decisions on key performance areas, allowing them to monitor, enable, enact, and evaluate risks. The people-side of innovations should not be forgotten: most tend to be involved on a part-time basis have high turnover rates and experience euphoria in the beginning, frustration and pain in the middle, and closure at the end of the innovation process.

The *implementation phase* is when the innovation is adopted and introduced. This includes “replicating and scaling up”, referring to ways to promote the rapid and effective expansion of an innovation in a public service organization. The implementation of an innovation occurs throughout the developmental period by linking and integrating the “new” with the “old” or by reinventing the innovation to fit the local situation (Rogers 2003). However, the role of the implementers of the innovation is often forgotten, even if implementers can use their discretion to adapt the innovation to specific circumstances during this process (Tummers et al. 2012).

The last phase of the innovation life cycle is the *evaluation phase*, consisting of analysis and learning from the innovation process. It requires a formal evaluation process to be established in order to identify what works and what does not, and so

promote continuous learning and improvement. It consists of the assessment of results in terms of output and outcomes. However, it goes a step further by integrating the innovation into the ongoing routine, and promoting it to others (Rogers 2003).

These phases strike a balance between a microassessment and macroassessment of innovation initiatives.

The evaluation phase can take place *ex-ante* (before deciding to innovate and implement the innovation), during the innovation life cycle, and *ex-post*. *Ex-ante* evaluation is aimed at investigating the potential and expected effects of innovation in the short, medium, and long term, thus defining a plan for its implementation (allocation of resources, responsibilities, and definition of activities). Evaluation during the innovation life cycle is aimed at monitoring the implementation of the innovation and at detecting whether the expected effects are being met or not. Finally, the purpose of *ex-post* evaluation is to define the effects obtained and foster the adoption of the innovation in case of positive results, promoting it as a best practice (Nasi 2013).

As is suggested in the following sections of this chapter, over the past four decades the public sector has also started looking into evaluating innovation performance, and a culture for measuring the impacts of innovation has started to develop. In particular, the evaluation culture in the public sector seems to be linked to the willingness to enhance transparency, learning, apprising, and proving accountability (Kattel et al. 2013). As Osborne and Gaebler (1992) say: “what gets measured, gets done”.

It is important to point out that although the public sector shares some parallel goals with the private sector (efficiency, productivity, and user satisfaction) and has therefore imported several performance instruments from the latter (such as ROI and other financial measures), the effects of an innovation in a PSO are more difficult to measure than those produced in a private business.

The main targets of operations in the private sector can be summed up with the concepts of profitability, market share, customer loyalty, and the economic nature of the business. Consequently, the private sector’s achievements can be measured by means of economic and financial measures (such as cost-benefit analysis, ROI, Net Present Value, Discounted Cash Flow, and Internal rate of Return).

On the other hand, the goals of the public sector are much more far-reaching: the public sector is concerned with protecting public interests, satisfying the various stakeholders, regaining confidence, and the value of public services. As a result, the targets of the public sector cannot be measured using economic and financial measures alone: they also require the use of non-financial and noneconomic indicators.

11.3 Research Methodology

11.3.1 *Systematic Review of Literature*

Once the purpose of the research and the framework of interest have been defined, the methodology adopted requires explanation. In order to answer the research question, a review of existing literature was completed by conducting a systematic review of literature.

Systematic literature review has been defined as “*a replicable, scientific and transparent process... that aims to minimize bias through exhaustive literature searches of published and unpublished studies and by providing an audit trail of the reviewer’s decisions, procedure and conclusions*” (Transfield et al. 2003).

Clear eligibility criteria must be established prior to embarking upon such a process.

11.3.1.1 Eligibility Criteria

Eligibility criteria for the study characteristics were determined for our systematic review in order to select the correct portion of existing literature. According to Liberati et al. (2009), such criteria can be listed as follows:

- Study design: empirical (e.g. case studies, experiments, and questionnaires) and theoretical studies;
- Year of publication: studies published between 1970 and 2013, since the first relevant literature on innovation performance was written in the 1970s or later;
- Language: only records written in English;
- Publication status: only international peer-reviewed journal articles.
- Type of studies—records should deal with innovation performance in the public sector, focusing in particular on the analysis of its effects and the measurement strategy adopted.
- Topic—records should contain these words in their entire text:
 - innovat*”
 - AND
 - “performance*” OR “evaluation*” OR “impact” OR “effect” OR “output” OR “result” OR “return on investment” OR “assessment” OR “outcome”

11.3.1.2 Search Strategy and Process

The systematic review was conducted on thirteen internationally preeminent journals, covering three macroareas: public management, managerial studies, and electronic government studies:

<i>FIELD OF STUDY</i>	<i>JOURNAL</i>
Public administration and management	Journal of Public Administration Research and Theory
	Public Administration
	Public Administration Review
	Public Management Review
	The American Review of Public Administration
	International Public Management Journal
Electronic government	Government Information Quarterly
	Electronic Journal of e-Government
	International Journal of Electronic Government Research
	Social Science Computer Review
	HICCS Proceedings (1995 - 2013)
Management studies	Journal of Management
	Journal of Management Studies

Fig. 11.2 Journals included in the systematic review

- Public administration journals (Journal of Public Administration Research and Theory, Public Administration Review, Public Administration, Public Management Review, American Review of Public Administration, International Public Management Journal);
- E-government journals (Government Information Quarterly, HICSS Proceedings, International Journal of E-Government Research, Electronic Journal of E-Government, Social Science Computer Review);
- Management journals (Journal of Management Studies, Journal of Management) (Fig. 11.2).

11.3.1.3 Study Selection

First, we applied the above-mentioned process to journals on Public Management and Management Studies. The resulting records were 263.

We then applied the same methods to Electronic Government Journals and found 114 records.

Summing up, the systematic literature review revealed 377 (263 + 114) records.

As far as our exclusion criteria are concerned, we decided that certain records were not pertinent after reading the articles themselves: we excluded articles that failed to fit the definitions of our streams of research and articles on the private sector.

After applying the above-mentioned exclusion criteria, we were able to ascertain that:

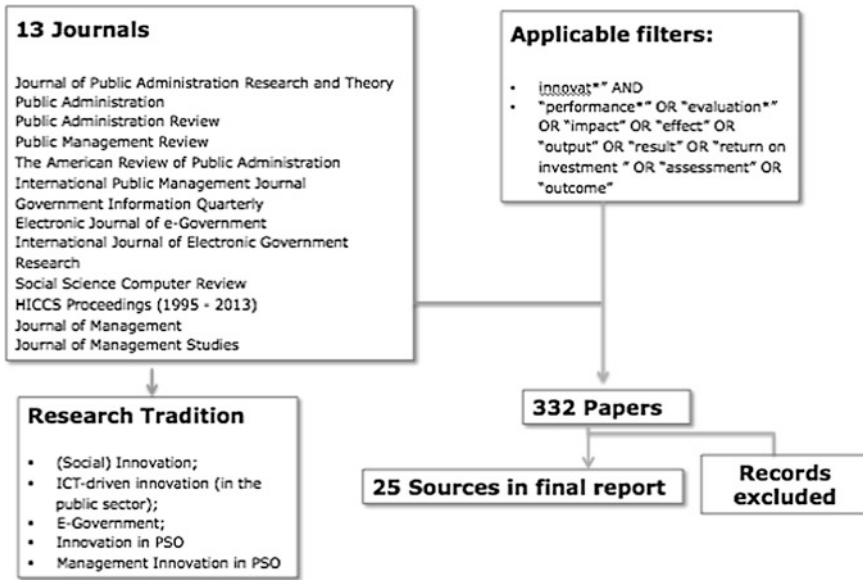


Fig. 11.3 The study selection process

- 78 records dealt with the private sector (Chang et al. 2013; Ciabuschi et al. 2011)
- 274 records did not deal with innovation performance:
 - 71 were related to the determinants of innovation (Walker 2008, 2013);
 - 203 were not relevant (e.g. innovation referred to the innovativeness of the research; innovation was not the main topic; the record names “Innovation Awards” in the text; the record deals with innovation awards…) (Fig. 11.3).

As a result, 25 records were found to be relevant to the topic of innovation performance in the public sector. These included:

- 20 in public management and managerial studies journals;
- 5 in e-government journals.

11.4 Findings

11.4.1 Study Characteristics

This section describes the characteristics of the 25 records found in our systematic literature review.

11.4.1.1 Journals

As mentioned above, the records included after our systematic review (25) are articles published in different peer-reviewed international journals focusing on public administration and management, electronic government, and management studies. The number of records included for each journal as a result of this process is illustrated below (Fig. 11.4).

11.4.1.2 Period of Publication

It is worth noting that the time of publication of the 25 records suggests increasing scientific interest in evaluating innovation performance in the public sector:

- 2 records were published in the period 1970–990;
- 2 records were published in the period 1991–2000;
- 10 records were published in the period 2001–2010;
- 11 records were published in the period 2011–2013.

The fact that most interest in innovation performance was noted in public management journals (17 total studies in the period of time we considered, with an increasing number of studies over time) is indicative of the paramount necessity of PSOs to monitor the effect of innovation. As mentioned earlier, assessing

FIELD OF STUDY	JOURNAL	NUMBER OF RECORDS INCLUDED
Public administration and management <i>(17 records included)</i>	Journal of Public Administration Research and Theory	3
	Public Administration	4
	Public Administration Review	2
	Public Management Review	4
	The American Review of Public Administration	3
	International Public Management Journal	1
Electronic government <i>(5 records included)</i>	Government Information Quarterly	2
	Electronic Journal of e-Government	0
	International Journal of Electronic Government Research	1
	Social Science Computer Review	2
	HICCS Proceedings (1995 - 2013)	0
Management studies <i>(3 records included)</i>	Journal of Management	1
	Journal of Management Studies	2
TOTAL NUMBER OF RECORDS INCLUDED		25

Fig. 11.4 Field of study and records included in the analysis

innovation adoption is a core phase in the innovation's life cycle (Tidd and Bessant 2011) especially in times of crisis: it enables the impacts of innovation to be measured in terms of economic and financial returns as well as non-monetary factors (key performance indicators of PSOs, such as the enhancement of working conditions and better quality of life for the patient).

11.4.1.3 Geographical Contexts

It has been seen that not all countries have developed a culture for measuring the impacts of innovation adoption (Bouckaert 2012) and so it would be useful to analyze the geographical distribution of the studies focusing on this issue, in order to investigate which countries are developing a culture that is more prone to evaluating the adoption of innovation in PSOs than others.

All the records resulting from the literature review were carried out in developed countries. In particular, Europe (UK, Denmark, Finland, Spain, Italy, and Germany) with 10 records; North America (USA) has 4 records; South America (Brazil) with 1 record; Central America (Mexico) with 1 record; Middle East (Israel) with 2 records; and Asia (South Korea, Hong Kong) with 2 records.

We found that the geographical distribution of the studies focusing on innovation evaluation is consistent with the need averted to measure the impacts of innovation in the different countries. In fact, greater attempts to measure and evaluate innovation were seen in South Korea with its Government Innovation Index (GGI), in Northern European countries with the project for Measuring Public Innovation in Nordic Countries (MEPIN), in the UK with the institutionalization of the National Endowment for Science Technology and Arts (NESTA) and the work undertaken by the National Audit Office (NAO) (Fig. 11.5).

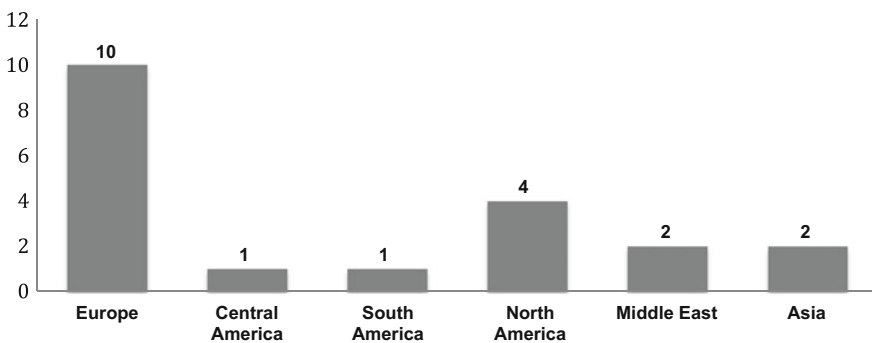


Fig. 11.5 Geographical distribution of studies

11.4.1.4 Research Methods

Additional relevant information emerged after observing the methodology used in these articles, with the most relevant facts summarized below:

- Most of the articles are empirical (16; 64%) (Andrews et al.2011; Damanpour et al. 2009), whereas a smaller proportion is theoretical (6; 24%) (Dewett and Jones 2001) and the remainder combines both typologies (3; 12%) (Vashdi et al. 2013) (Fig. 11.6).
- As regards the empirical studies, the majority (15; 79%) is quantitative (Walker et al. 2011; Damanpour et al. 1989), with only 3 (16%) qualitative works (Ahn and Bretschneider 2011). Only 1 study (5%) is based both on qualitative and on quantitative data (Cucciniello and Nasi 2013) (Fig. 11.7).

This information can be considered insightful with respect to the main research trends in this field: the net prevalence of empirical studies indicates the need averted to quantify the impacts of innovation in its different applications. However, as we will see later in the chapter, these analyses are unstructured, case-specific, and lack external validity. It follows that there is need for a common evaluation method structured in a multidimensional framework. Such an evaluation method would allow for comparisons between the impacts of innovation and so would lead to greater external validity of results.

Moreover, the majority of quantitative works rely on surveys (16; 84%) and some match survey data with census data or other existing datasets (Walker et al. 2010). It is also interesting that 2 studies are based on content analysis (Cheung 2005; Pope et al. 2006).

Finally, we noted that the majority of empirical studies in our literature review were conducted according to a panel or cross-sectional research design (Walker et al. 2011; Andrews et al. 2012; Damanpour et al.,1989). As Sillanpää (2013) pointed out, the measurement of innovation seems to lack measures capable of capturing the long-term effects and impacts of innovation. This difficulty is intrinsic to panel designs with few years of observation, and in cross-sectional design where many observations are captured for different subjects over the course of a single year.

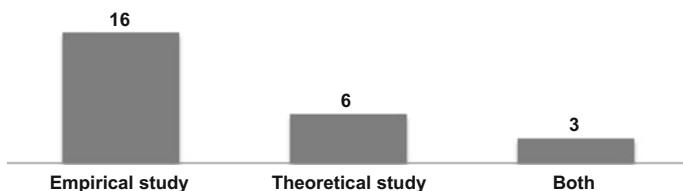
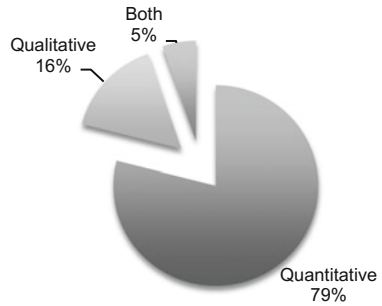


Fig. 11.6 Type of study

Fig. 11.7 Nature of the empirical studies



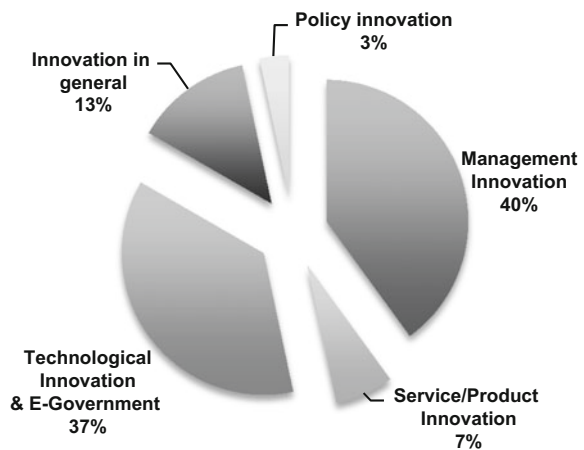
11.4.2 Categories of Innovation Considered

The following types of innovation are analyzed in the 25 records:

- Management innovation (12)
 - Managerial innovation (7)
 - Organizational innovation (2)
 - Administrative innovation (3)
- Service/product innovation (2)
- Innovation in general (e.g. innovative culture) (4)
- Policy innovation (1)
- Technological innovation (8) and e-government (3) (Fig. 11.8)

The total number of the types of innovation is not 25 because more than one type of innovation is considered in some records. Damanpour et al. (2009), for example, use the variable of “total innovation” which reflects the cumulative adoption of service innovation, technological innovation, and administrative innovation.

Fig. 11.8 Types of innovation



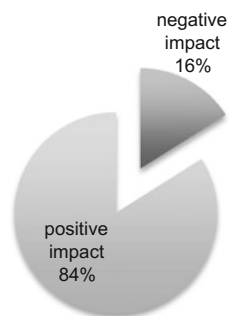
However, most studies deal with single innovations and not aggregate measures of innovation (Norris 1989; Cheung 2005; Cucciniello and Nasi 2005). Our analysis of the 25 records revealed that the evaluation of single types of innovation is often conducted using an unstructured and case-specific method, i.e. the authors have not developed a flexible measurement framework that could be applied easily to the same type of innovation in a different context.

11.4.3 Negative and Positive Effects

Before classifying the records in terms of output and outcome, it is interesting to note that 84% (21 records) of the records register positive aggregate effects of innovation, whereas the remaining 16% (4 records) found negative impacts of the innovation, although these negative effects could be attributed to the characteristics of the method used for measurement. Andrews et al. (2012), for example, found that innovative strategic stances result in overspending (because of the commitment to the development of new services). A potential disclaimer in this case could be related to the brief period of time considered, which may not be sufficient to allow for any positive effects of innovation to be measured. Norris (1989) noted another example of a negative effect, finding that local government employees using microcomputers reported stress and frustration at the beginning, but they also reported that their productivity increased once they had become proficient users of the PC.

There are several studies identifying positive effects. Yang and Kassert (2010), for example, found that managing by results, operationalized by way of perceived performance orientation and innovative culture, is positively related to job satisfaction, but the relationship is moderated by the confidence of employees in their leaders and their perceptions of the effectiveness and fairness of performance appraisal. Vigoda et al. (2008) found that public sector innovation (entrepreneurial actions, creativity, flexibility, and willingness to adopt new ideas) has an impact on the confidence and satisfaction with public administration and that this effect is both direct and mediated by the image of public organizations (Fig. 11.9).

Fig. 11.9 Negative vs. positive effect



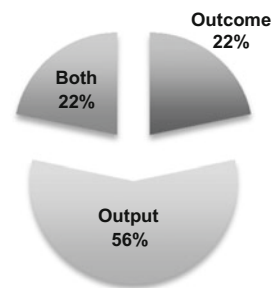
11.4.4 *Outputs and Outcomes*

It may be useful to define the terms “outcome” and “output” in order to indicate how the initial classification of the results for the 25 records may be carried out.

Most discussions of performance measurement are based on an implicit model of the production process, in which inputs are used to create outputs. Outputs result in outcomes, which have also been described as the consequences, results, effects, or impacts of service provision. According to Boyne and Law (2005), outcomes can be conceptualized in a number of ways. First, they can be divided into those that measure intermediate or final outcomes. Final outcomes are the ultimate consequences of the outputs produced by public organizations and refer to the achievement of the purposes of public organizations (such as better health care or fewer crimes) and focus on the recipients of a service rather than the characteristics of the service itself. Intermediate outcomes are the result of service provision (for example, the number of people who stop smoking or the number of offenders appearing in court) and represent a step on the path to a final outcome (Hatry 2001).

According to the above definition of outcomes and outputs, we found that the majority (13; 56%) of our studies deals with the analysis of outputs. Andrews et al. (2012) analyze the effect of innovation on overspending in local governments; Lee and Perry (2002) examine whether investments in information technology lead to an increase in aggregate performance measured as the gross state product. A quarter of the studies (5 records; 22%) focus on outcomes, such as Jaeger (2005) who investigated the effects of e-Government on democratic dialogue. Finally, an identical proportion of studies (5; 22%) deal with both outcomes and outputs: Moynihan (2003) examines how innovative participation technologies in local administrations reduce administrative costs and raise instrumental benefits, reinvigorating public hearing (Fig. 11.10).

Fig. 11.10 Outputs vs. outcomes



11.4.5 Dimensions of Performance

11.4.5.1 Innovation Performance Dimensions and Indicators in the Public Sector

As stated in the introduction, even though the private and the public sectors share some parallel goals, the targets of the public sector cannot be measured using economic and financial measures alone: they also require the use of non-financial and noneconomic indicators. This fundamental difference is because the private sector is mainly concerned with issues related to profitability, whereas the public sector cares about protecting public interests, satisfying its different stakeholders, recovering confidence, and enhancing the value of public services.

There are three studies in the records we selected in our literature review that suggest adopting non-financial and noneconomic indicators adapted from the private sector: Germbergen and Amenlickx (2002) with the Balanced Score Card, Walker et al. (2002) with the Literature-based Innovation Output Indicator (LBIOI), and Millar and Hall (2013) with the Social Return on Investment (SROI).

The Balanced Scorecard (Kaplan and Norton 1996) was conceived as a simple performance measurement framework and is a full strategic planning and management system today. It classifies the strategic goals of organizations according to four main dimensions: financial/stewardship (financial performance), customer/stakeholder (satisfaction), organizational capacity (knowledge and innovation), and internal business processes (efficiency). On the other hand, the SROI method is designed to understand, manage, and report the social, environmental, and economic value created by an organization (New Economics Foundation 2004). This method has been employed primarily in the UK, where policy makers have encouraged its adoption in social enterprises, especially in the Department of Health. Finally, the LBIOI originated in the work of Edwards and Gordon (1984), and Kleiknecht (1993). LBIOs were originally generated by sampling new product announcement sections belonging to the technical and trade sectors; they have only recently been applied to the public sector to track reported innovations over time in order to explore their adoption, their diffusion rate, and their impacts.

The latter three mentioned records and Cucciniello and Nasi (2013) are the only studies in our literature review that propose a theoretical framework to evaluate innovation performance. The other records display unstructured and case-specific evaluation methodologies. An example of an unstructured and case-specific characteristic is found in Cheung (2005): in this study, the author examines the capacity of performance pledges in Hong Kong in informing and empowering customers of public services with respect to access, choice, information, and representation, applying the five-principle framework for public sector consumerism. However, the application of this framework does not lead to an extensive examination of the effects of performance pledges: it focuses only on the consumerist aspect and only considers one of the several dimensions of impacts linked to performance pledges.

In general, the records considered do not propose a theoretical evaluation method and the resulting application of the method on the innovation of interest: they analyze the effects of an innovation already applied (only ex-post evaluations are carried out) with respect to the dimensions the authors prefer to investigate (efficiency, or effectiveness, or public value). There is no study that carries out an examination including all of these dimensions and so the methods proposed cannot be generalized or applied to other innovations, even if they are the same type.

A disclaimer in favour of the innovation evaluation methodologies adopted in the records is that none of them, other than the four mentioned above, aims to propose a general evaluation framework or externally valid results. Apparently, all the literature in the journals of interest aims to assess ex-post evaluations of specific innovations that take place in specific internal and external environments.

11.5 Some Considerations on the Results of Our Systematic Literature Review

The problems of innovation evaluation in the public sector that emerged in our literature review can be summed up as follows:

First, the concept of innovation itself seems to be abused in literature: any seemingly significant change in public service delivery, in the culture of the PSO, or in the organizational setup, is considered to be an innovation. It is important to point this out because this conceptual incoherence could make it impossible to evaluate the impacts of innovation. For example, some records focus on the concepts of innovative culture and innovativeness (Vigoda-Gadot and Meiri 2008), but these notions of innovation appear too far reaching for any results characterized by external validity to be deduced.

Second, measuring performance in the public sector is clearly a complex process and brings many difficulties with it. The dimension of efficiency, and of effectiveness to a lesser extent, (they will be extensively analyzed in the following sections) seems to be preferable options when measuring innovation performance, as they enable “easily” quantifiable measures to be obtained that can promote the subsequent adoption of innovation within the PSO. However, as stated above, the public sector should not employ indicators that are only limited to the dimension of efficiency, but indicators capable of assessing the degree of satisfaction of the interests of all the stakeholders involved. For instance, Lee and Perry (2002) considered the effect of investments in ITC with respect to the efficiency dimension alone by looking at their impacts in terms of productivity.

Third, the evaluation methods employed in the records under scrutiny are developed based on case-specific, unstructured frameworks and so they cannot be applied to other contexts. This undermines the external validity of the results and threatens the possibility to generalize them and develop policy directives or guidelines.

Fourth, the measurement of innovation performance is often conceived as static: the majority of the studies considered adopt cross-sectional or panel frameworks with a limited number of years of observations, or the period of time between the adoption of innovation and the measurement of its impacts is too short, or no long-term view is taken into consideration. Some studies point out the necessity to consider an appropriate period of time in order to estimate the effects of innovation on performance in the public sector correctly (Sillanpää 2013; Damanpour et al. 2009; Wischnevsky and Damanpour 2006). The underlying intuition is that the potential performance benefits of organizational transformation may not materialize immediately, which may be due to the outlay and disruption associated with the implementation of innovation. In particular, Damanpour and Evan (1984) found that the degree of organizational lag is inversely related to organizational performance, and other studies therefore employed periods of time ranging from one year (Damanpour et al. 2009) to seven years (Damanpour 1990) in order to assess the impacts of innovation. Finally, in some cases the success of the innovation can only be ascertained in terms of long-term effectiveness, which can usually only be assessed through qualitative measurements.

By way of conclusion, the problem of defining indicators in the public sector is not only technical; it is also conceptual, as stated by Van de Walle (2008).

The lack of a structured methodology could be due to the limited diffusion of evaluation frameworks capable of benchmarking innovative practices and measuring their opportunity costs. Williams (2001) pointed out that evidence portrayed in the form of impacts delivered after the adoption of an innovation remains methodologically limited.

Of all the records we investigated, only the work by Cucciniello and Nasi (2013) presented a comprehensive and structured method of evaluation. The authors constructed a multidimensional evaluation framework that can support the assessment of improvements in organizational performance in a flexible and user-friendly manner: in order to assess the impacts of Electronic Medical Records (EMRs), they considered four main dimensions of impact: efficiency, organizational effectiveness, clinical governance, and quality of supporting services. Each dimension was then divided into fifteen impact categories and a total of forty-one indicators. These dimensions and their corresponding measures enabled the evaluation of the effects of EMRs on the performance of the health care organization by considering them disjointedly, since they affect different spheres of the health care organization, its community, and the local territory.

A former antecedent of this methodological framework is represented by the eGovernment Economics Project (eGEP), which was built around three value drivers: efficiency, democracy, and effectiveness, and was designed to produce a multidimensional assessment of the public value potentially generated by eGovernment.

By way of conclusion, previous research has proven the need for a comprehensive method for measurement that allows for the evaluation of quality-related effects through non-financial/noneconomic indicators and also for the assessment of long-term effects. In relation to the former issue, the latter two models (Cucciniello

and Nasi 2013) may represent a significant model for setting the road map for evaluating innovation performance. However, more scientific research is required as regards the time period to be considered when evaluating the impacts of innovation, and it would be interesting to understand the period that needs to be considered when evaluating different types of innovation (i.e. administrative, technological, and managerial).

11.6 Discussion

Currently available literature on innovation in the public sector seems to suggest the following interconnected trends:

- Evidence collected from literature as to the effects of innovation is scant and no quantitative analysis, such as meta-analysis, helps to assess which type of innovation has a significant impact on which dimension of performance (efficiency, effectiveness, return on investment, public value) at macrolevel.
- The ambiguousness of evidence is linked to the lack of structured theoretical methods capable of assessing and measuring the impacts of innovation from different perspectives (such as strategic value, public value, or economic/financial value). As a result, the measurement techniques that are adopted are not comparable and rely on single measures of performance, threatening the external validity of results.
- The main barriers to the measurement and evaluation of innovation are represented by the following factors: first, there is a lack of a solid evaluation culture and many government evaluation frameworks have spread during the last decade, such as South Korea's "Government Innovation Index (GII)" project, the "Measuring Public Innovation in Nordic Countries (MEPIN)" project; NESTA's "Public Sector Innovation Index" project in the United Kingdom; the OECD's "National Experts for Science and Technology Indicators" (NESTI); the European Commission's "European Public Sector Innovation Scoreboard" (EPSIS) project, and the Australian Public Sector Innovation Index project (ASII), and so it seems that the initiative has not been assimilated in a mature way in the academic field. Second, there is lack of a commonly recognized definition of costs and benefits. Third, the studies considered do not specify who is the target of the evaluation, even though defining the target of the evaluation is crucial if we are to identify the proper methods of evaluation in terms of the dimensions, variables and measures to consider. Moreover, since public sector activity aims to be accountable to several categories of stakeholder (such as the general public, politicians, businesses, and NGOs), identifying the target of interest would be useful for setting the purpose of the evaluation.

11.7 Setting the Way Forward

To conclude our analysis, we would like to stress the relevance of measuring and evaluating innovation in today's context.

The positive contributions made by innovation in the public sector are widely recognized in both the academic and non-academic worlds, as demonstrated by the appearance of the word "innovation" in every politician's agenda. In particular, the main benefits attributed to innovation have been summarized by the following contributions: the delivery of "better services" (Osborne and Brown 2005); improved quality of life of individuals and territories; enhanced government performance enhanced competitiveness (Thenint 2010; Setnikar Cankar et al. 2013); and maintenance and enhancement of "trustworthy relationships" with constituents (Mulgan and Albury 2003).

Innovation may enhance organizational performance and may improve competitiveness and the quality of life of territories and individuals.

As a result, further studies would be useful in order to address several key issues:

- There seems to be overuse in the labelling of any seemingly relevant change in a PSO, in public service delivery, or organizational setup as an innovation.
- Measuring performance in the public sector is a complex and complicated matter and there seems to be a lack of a general and comprehensive framework, allowing for external validity of results.
- The period of time considered for detecting the effects of innovation is not always appropriate (short run perspective).

Further studies using experimental approaches, based on comparisons involving different countries, may help us to address these issues and develop a culture that recognizes the evaluation of innovation as a prerequisite for managing innovation in the public sector.

The term "innovation life cycle" should be linked to evaluation as a continuous process that helps us decide whether to continue to innovate, how to intervene in order to get the expected results in a more efficient and effective way, how to meet stakeholder expectations, as well as helping to continuously improve the evaluation process.

References

- Ahn, M. J., & Bretschneider, S. (2011). Politics of E-government: E-government and the political control of bureaucracy. *Public Administration Review*, 71(3), 414–424.
- Andrews, R., Boyne, G. A., & Walker, R. M. (2012). Overspending in public organizations: does strategic management matter? *International Public Management Journal*, 15(1), 39–61.
- Bason, C. (2010). *Leading public sector innovation*. Bristol: Policy Press.

- Bloch, C. (2010). *Towards a conceptual framework for measuring public sector innovation, module 1—Conceptual framework*. Aarhus: The Danish Centre for Studies in Research and Research Policy.
- Bouckaert, G. (2012). Public sector performance: Managing governments by the numbers. In G. Tria & G. Valotti (Eds.), *Reforming the public sector: How to achieve better transparency, service, and leadership*. Washington, DC: Brookings Institution Press.
- Boyne, G. A., & Law, J. (2005). Setting public service outcome targets: Lessons from local public service agreements. *Public Money & Management*, 25(4), 253–260.
- Chang, S., Gong, Y., Way, S. A., & Jia, L. (2013). Flexibility-oriented HRM systems, absorptive capacity, and market responsiveness and firm innovativeness. *Journal of Management*, 39(7), 1924–1951.
- Cheung, A. B. L. (2005). What's in a pamphlet? Shortfalls and paradoxical flaws in Hong Kong's performance pledges. *Public Management Review*, 7(3), 341–366.
- Ciabuschi, F., Dellestrand, H., & Martin, O. M. (2011). Internal embeddedness, headquarters involvement, and innovation importance in multinational enterprises. *Journal of Management Studies*, 48(7), 1612–1639.
- Cucciniello, M., & Nasi, G. (2013). Evaluation of the impacts of innovation in the health care sector: A comparative analysis. *Public Management Review*, 16(1), 90–116.
- Damanpour, F. (1990). Innovation effectiveness, adoption and organizational performance. In M. A. West & J. L. Farr (Eds.), *Innovation and creativity at work* (pp. 125–141). New York: Wiley.
- Damanpour, F., & Evan, W. M. (1984). Organizational innovation and performance: The problem of “organizational lag. *Administrative Science Quarterly*, 29(3), 392–409.
- Damanpour, F., & Schneider, M. (2009). Characteristics of innovation and innovation adoption in public organizations: Assessing the role of managers. *Journal of Public Administration Research and Theory*, 19(3), 495–522.
- Damanpour, F., Szabat, A., & Evan, W. (1989). The relationship between types of innovation and organizational performance. *Journal of Management Studies*, 26(6), 587–602.
- Damanpour, F., Walker, R. M., & Avellaneda, C. N. (2009). Combinative effects of innovation types and organizational performance: A longitudinal study of service organizations. *Journal of Management Studies*, 46(4), 650–675.
- Dawes, S. S., Cresswell, A. M., & Pardo, T. A. (2009). From “need to know” to “need to share”: Tangled problems, information boundaries, and the building of public sector knowledge networks. *Public Administration Review*, 69(3), 392–402.
- Dewett, T., & Jones, G. R. (2001). The role of information technology in the organization: A review, model, and assessment. *Journal of Management*, 27(3), 313–346.
- Dunleavy, P., Margetts, H., Bastow, S., & Tinkøer, J. (2009). New public management is dead—Long live digital-era governance. *Journal of Public Administration Research and Theory*, 16, 467–494.
- Edwards, K. L., & Gordon, T. J. (1984). *Characterization of innovations introduced in US market since 1982*. Glastonbury, CT: Futures Group.
- Fagerberg, J., Mowery, D. C., & Nelson, R. R. (2005). *The Oxford handbook of innovation*. Oxford: Oxford University Press.
- Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of innovations in service organizations: Systematic review and recommendations. *Milbank Quarterly*, 4, 29–581.
- Jaeger, P. T. (2005). Deliberative democracy and the conceptual foundations of electronic government. *Government Information Quarterly*, 22(4), 702–719.
- Kaplan, R. S., & Norton, D. P. (1996). Using the balanced scorecard as a strategic management system. *Harvard Business Review*, 74(1), 75–85.
- Kattel R., Capilovs A., Drechsler W., Kalvet T., Lember V., & Tonurist P. (2013). *Can we measure public sector innovation? A literature review*: Lipse Project Paper.
- Kim, Y., Lee, Z., & Gosain, S. (2005). Impediments to successful ERP implementation process. *Business Process Management Journal*, 11(2), 158–170.

- Kleiknecht, A. (1993). *Towards a literature-based innovation indicators*. Amsterdam: SEO.
- Lapsley, I., & Llewellyn, S. (1998). Markets, hierarchies and choices in social care. *Management*, 4, 29–38.
- Lee, G., & Perry, J. L. (2002). Are computers boosting productivity? A test of the paradox in state governments. *Journal of Public Management Research and Theory*, 12(1), 77–102.
- Liberati, A., Altman, D. G., & Teztlaff, J. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Annals of Internal Medicine*, 151(4), 65–94.
- Millar, R., & Hall, K. (2013). Social return on investment (SROI) and performance measurement. The opportunities and barriers for social enterprises in health and social care. *Public Management Review*, 15(6), 923–941.
- Moore, M. H. (1994). Public value as the focus of strategy. *Australian Journal of Public Administration*, 53, 296–303. doi:10.1111/j.1467-8500.1994.tb01467.
- Moore, M. H. (1995). *Creating public value: Strategic management in government*. Harvard university press.
- Moynihan, D. P. (2003). Normative and instrumental perspectives on public participation: Citizen summits in Washington, D.C. *The American Review of Public Administration*, 33(2), 164–188.
- Mulgan G. & Albury D. (2003). *Innovation in the public sector: Strategy unit*. Cabinet Office, 1–40.
- Nasi, G. (2013). *Misurare e valutare l'innovazione nelle aziende pubbliche*. Milan: Egea.
- New Economics Foundation (NEF). (2004). *Social return on investment: Valuing what matters*. London: New Economics Foundation.
- Norris, T. (1989). High tech in city hall: Uses and effects of microcomputers in United States local governments. *Social Science Computer Review*, 7(2), 242–244.
- Osborne, D., & Gaebler, T. (1992). *Reinventing government*. Lexington, MA: Addison-Wesley.
- Osborne, S. P., & Brown, K. (2005). *Managing change and innovation in public service organizations*. London: Routledge.
- Pope, C., Robert, G., Bate, P., Le May, A., & Gabbay, J. (2006). Lost in translation: A multi-level case study of the metamorphosis of meanings and action in public sector organizational innovation. *Public Administration*, 84(1), 59–79.
- President Obama. (2013). *Memorandum to the heads of departments and agencies*. <http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-17.pdf>.
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York: Free Press.
- Setnikar-Cankara, S., & Petkovšek, V. (2013). *The health care system in Slovenia. Health reforms in Central and Eastern Europe: Options, obstacles, limited outcomes* (pp. 221–234). The Hague: Elen.
- Sillanpää, V. (2013). Measuring the impacts of welfare service innovations. *International Journal of Productivity and Performance Management*, 62(5), 474–489.
- Stefanou, C. J. (2001). A framework for the ex-ante evaluation of ERP software. *European Journal of Information Systems*, 10(4), 21–595.
- Thenint, H. (2010). *Mini Study 10: Innovation in the public sector: Global review of innovation intelligence and policy studies*. Inno Grips.
- Tidd, J., & Bessant, J. (2011). *Managing innovation: Integrating technological, market and organizational change*. New York, NY: Wiley.
- Tranfield, D., et al. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, 14(3), 207–222.
- Tummers, L., Vermeeren, B., Steijn, B., & Bekkers, V. (2012). Public professionals and policy implementation: Conceptualizing and measuring three types of role conflicts. *Public Management Review*, 14, 1041–1059.
- Van de Walle, S. (2008). Comparing the performance of national public sectors: Conceptual problems. *International Journal of Productivity and Performance Management*, 57(4), 329–338.

- Vashdi, D. R., Vigoda-Gadot, E., & Shlomi, D. (2013). Assessing performance: The impact of organizational climates and politics on public schools' performance. *Public Administration, 91*(1), 135–158.
- Vigoda-Gadot, E., & Meiri, S. (2008). New public management values and person-organization fit: A socio-psychological approach and empirical examination among public sector personnel. *Public Administration, 86*(1), 111–131.
- Vigoda-Gadot, E., Shoham, A., Schwabsky, N., & Ruvio, A. (2008). Public sector innovation for Europe: A multinational eight-country exploration of citizens' perspectives. *Public Administration, 86*(2), 307–329.
- Von Hippel, E. (2006). *Democratizing innovation*. Cambridge, MA: MIT Press.
- Walker, R. M. (2008). An empirical evaluation of innovation types and organizational and environmental characteristics: Towards a configuration framework. *Journal of Public Administration Research and Theory, 18*, 591–615.
- Walker, R. M. (2013). Internal and external antecedents of process innovation: A review and extension. *Public Management Review, 16*(1), 21–44.
- Walker, R. M., Boyne, G. A., & Brewer, G. A. (Eds.). (2010). *Public management and performance: Research directions*. New York: Cambridge University Press.
- Walker, R. M., Damanpour, F., & Devece, C. A. (2011). Management innovation and organizational performance: The mediating effect of performance management. *Journal of Public Management Research and Theory, 21*(2), 367–386.
- Walker, R. M., Jeanes, E., & Rowlands, R. (2002). Measuring innovation—Applying the literature-based innovation output indicator to public services. *Public Administration, 80*(1), 467–494.
- Wischnevsky, J. D., & Damanpour, F. (2006). Organizational transformation and performance: An examination of three perspectives. *Journal of Managerial Issues, 18*(1), 104–128.
- Yang, K., & Kassert, A. (2010). Linking management reform with employee job satisfaction: Evidence from federal agencies. *Journal of Public Management Research and Theory, 20*(2), 413–436.

Chapter 12

Measuring the Impact of Training on Public Manager Performance: The Case of Canton Ticino (An Ex-Post Analysis)

Andrea Martone, Filippo Sciaroni and Alan Righetti

Abstract The importance of training as a determiner of a company's performance is widely acknowledged: the "Canton Ticino" (Switzerland) has to demonstrate that the organisational decisions (such as whether or not to maintain a training programme) are based on a strategic learning process. The aim of this chapter is to illustrate the evaluation process of a training programme conducted for a group of Swiss public managers. In order to measure the impact of the training, the first three (Reaction, Learning, Behaviour) of the four levels identified in Kirkpatrick's model (1994) were measured, using an ex-post analysis. In particular, the chapter is focused on the changes at behavioural level.

Keywords Training evaluation · Public management · Public value creation · Public performance measurement

A. Martone (✉) · F. Sciaroni · A. Righetti
SUPSI, Lugano, Switzerland
e-mail: andrea.martone@supsi.ch

F. Sciaroni
e-mail: filippo.sciaroni@supsi.ch

A. Righetti
e-mail: alan.righetti@supsi.ch

12.1 Introduction

In 2011, the Ticino Public Administration decided to launch a major training programme with the aim of preparing its cantonal management (staff classified as Public Manager—*Funzionario-Dirigente*¹-FD) to confront the new challenges that the Ticino public sector will have to face in the coming years.

In effect, the socio-economic framework in which the cantonal administration authority will operate appears rather complex due to the crisis in the Swiss banking-financial system that has set two completely new challenges for all the public administrative authorities:

- Financial crises (the Ticino Administrations' fiscal revenues have decreased significantly, resulting in a balance sheet deficit that the administrative authorities were immediately called upon to respond to by cutting costs and increasing tax rates);
- Rethinking of the cantonal economy, which can no longer count on the traditional employment *drivers* (banks, trust companies and financial consultancies); therefore, in order to ensure future generations the same level of wellbeing as that enjoyed by previous ones, a new development model has to be envisaged for the region.

In this context, the role of the Canton changes. Instead of being an efficient manager of public assets, it has to become a promoter of socio-economic change; a sphere where, in addition to political management, also technical management should play a different role that is more active and more entrepreneurial. As a consequence, a training programme entitled “*Condurre e dirigere*” (*Lead and Manage*) was launched for all FDs (271 people), consisting of 20 classroom-based lessons that were spread over a 3 year period.

To define the content and the structure of the training programme, the administration utilised a specialist internal office, the USM (*Ufficio per lo Sviluppo Manageriale*²—Office for Managerial Development), and collaborated with the local *University of Applied Science* (SUPSI).

The design model of the training programme was that of *co-design*, which envisages extensive involvement of the trainees (FDs) not only in the

¹Article 24 of the LORD (Legge sull'ordinamento degli impiegati dello Stato e dei docenti del Canton Ticino—Law on the regulations of State employees and teachers of Canton Ticino) (1995) describes the role of Public Manager as follows: Public Managers organise, direct, coordinate and check the work of their collaborators.

They contribute to the promotion and implementation of all the measures aimed at improving the efficiency and the quality of the performance of their service, ensuring that the service operates correctly.

²USM is the current name of the office specialised in management training. It was originally called CEFOS (*CEntro FOrmazione e Sviluppo*—Centre for Training and Development) and this former name is used in some documents cited in this chapter.

information-gathering phase, but also during the design phase of the training and, subsequently, when verifying the results.

Moving on from this, in order to locate the training programme within the institutional framework of Ticino (which is not very well known outside the Cantons) a brief description of the Cantonal Administration is now given, followed by an analysis of the co-design process that was adopted to define the programme.

12.1.1 Cantonal Administration Authorities

“The Cantons are sovereign, except to the extent that their sovereignty is limited by the Federal Constitution. They exercise all rights that are not vested in the Confederation” (Art. 3 of the Federal Constitution 1999). Each canton has a constitution, a parliament, a government and autonomous legislating bodies; all areas of competence not specifically allocated to the Swiss Confederation are exercised by the Cantons, based on the principle of subsidiarity (*Art. 5, Cost, 1999*).

“Canton Ticino is a democratic republic of Italian culture and language” (Article 1 of the Constitution of the Republic and Canton of Ticino, 1997), part of the Swiss Confederation; the Ticinese people are *“faithful to the historic task to interpret Italian culture within the Helvetic Confederation” (Preamble of the Constitution of the Republic and Canton of Ticino, 1997).*

The Gran Consiglio (Parliament) is the legislative authority of Canton Ticino and consists of 90 members elected by the people every 4 years by a proportional system. *“The main activity of the Parliament is to legislate: as notified by one of its Commissions, the Gran Consiglio adopts, modifies or rejects the draft laws and legislative decrees submitted to it by the Consiglio di Stato – CdS (Council of State), put forward by the people, the Municipalities or the members of parliament” (Canton Ticino website 2016a).*

The executive power lies with the *“Consiglio di Stato, the governmental and executive authority of the Canton and consists of five members directly elected by the people [...] every four years, at the same time as the election of the Gran Consiglio” (Canton Ticino website 2016b).*

The Consiglio di Stato directs cantonal matters in a collegial manner, organising and performing its activities through five departments (Department of the Institutions; Department of Healthcare and Social Affairs; Department of Education, Culture and Sport; Department of the Territory; Department of Finance and the Economy), and other subordinate offices (see Fig. 12.1). Each member of the Consiglio di Stato is the Head of a Department, decided upon at the beginning of each legislature.

The President and the Vice President, who have representational roles, are appointed in rotation and remain in office for 1 year.

Figure 12.1 shows the organisational structure (first-level bodies) of the Cantonal Administration.

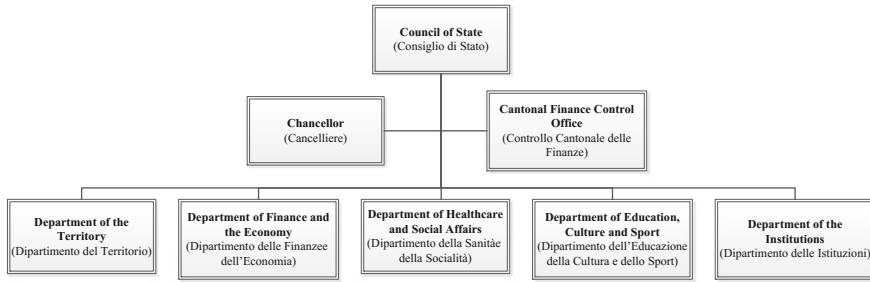


Fig. 12.1 Structure of the Cantonal Administration. *Source:* drawn up by the authors

The State Chancellor, who “*broadly speaking, carries out general staff functions in the fields of planning, organisation, preparation, coordination and checking*”, “*attends Consiglio di Stato meetings with an advisory vote*”, and “*coordinates the work between the Executive and Legislative areas*” (Canton Ticino website 2016c). In short, it may be said that s/he is the guarantor of the administrative propriety of State proceedings.

The Cantonal Finance Control Office is “*the financial control body of Canton Ticino. It checks the State accounts and balance sheet annually, and ensures that the Cantonal Administration services are audited in adherence to the Law on the financial control and management of the State (LGF)*” (Canton Ticino website 2016d).

With regard to the training programme, the courses were divided into three areas:

- CEFOS (previously named USM) courses: these are listed courses on various topics, unrelated to each other, which are offered to cantonal employees who enrol on a voluntary basis (after having received authorisation from their direct managers);
- Training projects and ad hoc courses, corresponding to tailor-made training programmes, which are organised only once, or perhaps in a few editions aimed at meeting specific training requirements that are expressed by specific organisational units.
- Compulsory training courses for public managers—the subject of the present chapter.

The data in Fig. 12.2 require some explanation: the number of participants (84 in 2013, 239 in 2014 and 361 in 2015) refer to the participants in the various courses; therefore, if a specific FD participated in more than one courses (as happens regularly every year), s/he was counted twice.

Figure 12.3 shows the official data that, in the 3-year period examined, the courses for FDs were short in duration (1.4 days on average), with an average attendance of 10.5 FDs per class (in line with the Cantonal targets).

	Courses				Participants				Course days			
	2012	2013	2014	2015	2012	2013	2014	2015	2012	2013	2014	2015
CEFO5 courses	18	27	25	25	283	355	367	288	31.0	41.5	32.0	23.0
Training projects and ad hoc courses	17	7	7	12	175	71	207	210	14.0	4.0	30.5	32.5
Compulsory training programs for public managers	0	8	19	30	0	84	239	361	0.0	16.0	31.0	50.0
Total	35	42	51	67	458	510	813	859	45	61.5	93.5	105.5

Fig. 12.2 Training courses. Source: Consiglio di Stato (2015)

Course participants and no. of lesson days.	2013	2014	2015	Average
Courses	8	19	39	
Participants	84	239	361	10.4
Lesson days	16	31	50	1.5

Fig. 12.3 Course participants and no. of lesson days. Source: Consiglio di Stato (2015)

12.1.2 Shared Design Process

A 3-phase procedure was followed in order to analyse the training needs and co-design the training programme for FDs:

1. Exploration: 83 short interviews (of 30 min) were conducted “in the field”, in order to understand the main work-related problems and the expectations related to the Cantonal Administration FD training programme;
2. In-depth analysis: completed by means of 40 long interviews (of 2 h), where the issues and problems that had emerged in phase one were re-discussed and analysed with the interviewees;
3. Design: based on the previous two phases, a training programme was drawn up, and subsequently submitted for the assessment of 5 focus groups, one for each department. A “technical focus group” was subsequently set up, composed of Human Resource (*Sezione Risorse Umane*—SRU) staff members. The new FD training programme was mapped out at the end of this phase.

The training programme resulting from this “shared design” process was structured into 13 courses that confronted issues related to the development of new managerial skills. When the Canton approved the courses, it also deliberated (RG 635 of 21/11/2012) that there would be an assessment of the impact made by the courses on the managerial skills of its managers in order to identify, based on Kirkpatrick’s model (1994)³, three training impact measurement levels⁴:

³Many authors refer to the training assessment: Training Validation System (TVS) by Fitz-Enz (1994); Input, Process, Output/ Outcome (IPO) by Bushnell (1990); Context, Input, Process, Product (CIPP) by Worthen and Sanders (1987); the five level evaluation model of Kaufman et al. (1995); the Context, Inputs, Reactions and Outcomes (CIRO) Approach of Warr et al. (1970). Holton (1998) focuses on the transfer process; Tannenbaum et al. (1991) on training effectiveness; Matthieu et al. (1993) on training outcomes. Finally, Noe (1986), Olsen (1998), Winfred and Winston (2003) and Van Buren and Erskine (2002) confirm the Kirkpatrick taxonomy effectiveness.

⁴Kirkpatrick’s 4th level has been excluded since it is extremely complex to record.

- Participant satisfaction;
- Content learning;
- Change in manager behaviour.

This model has been criticised by many authors (Alliger and Janak 1989; Alliger et al. 1997; Cannon-Bowers et al. 1995; Ford and Kraiger 1995; Salas and Cannon-Bowers 2001; ...). The criticism focused on three problematic assumptions: “(1) The levels are arranged in ascending order of information provided, (2) The levels are causally linked, and (3) The levels are positively inter-correlated” (Alliger and Janak 1989, p. 331). Nevertheless, the model was adopted by the Canton.

The aim of the present chapter was to illustrate the evaluation process of a training programme conducted for a group of Swiss PA managers and its impact, with particular reference to the third level (behaviour) of Kirkpatrick’s model.

12.2 Fundamental Theories and Reference Literature

The literature underlying the training programme may be traced back mainly to the authors who have previously discussed the topic of skills (Rullani 2004; Nonaka and Takeuchi 1995; Prahalad and Hamel 1990; Boyatzis 1982; Polanyi 1962). The programme was created precisely in order to develop the managerial skills of the FDs.

More specifically, four critical skills were identified:

- Tendency to public entrepreneurship;
- Tendency to efficiency (organisation and utilisation of managerial tools);
- Human resource management skills;
- Communication skills.

These are not the product of any technical assessment made by the administration or by the university, on the contrary they stem from suggestions made by the FDs in the course of the co-design process. The skills emerging from this shared design process were subsequently systematised within the reference literature, as will now be explained.

12.2.1 *Public Entrepreneurship Orientation*

The orientation of entrepreneurship may be defined as “an individual’s ability to turn ideas into action. It includes creativity, innovation and risk-taking, as well as the ability to plan and manage projects in order to achieve objectives. [...] This should include awareness of ethical values and promote good governance.” (European Union 2006, p. 17).

The European Union considers entrepreneurship as one of the eight “key competences” in Europe (European Union 2006, p. 13).

In the training programme, the concept of public entrepreneurship was divided into two components:

- Creation of public value, understood as the ability of the public manager to interpret the role of value creator for the reference territory. As cultural references, there was taken into consideration the theory of public value (Moore 1995; Meynhardt 2009) and the concept of strategic management of services (Norman 1992; Rebora and Meneguzzo 1990);
- Knowledge of public law, based on the study of current legislation and of the cantonal and federal body of laws.

12.2.2 Organisational Efficiency Orientation

The concept of organisational efficiency refers to the quantity of human resources used in order to achieve an organisational objective, where effectiveness is the extent to which the organisation achieves a stated objective (Daft 2015; Ansoff and Brandenburg 1971 et al.). For the Canton, efficiency was understood as a FD’s tendency of to optimise organisational structures and processes, and to make careful use of financial and technological resources. In terms of skills, this concept was divided into two components:

- Knowledge of managerial techniques: particularly organisation design, project management, budgeting, knowledge management, etc.
- Knowledge of specific work instruments, such as: techniques for time management, decision making, speed reading, mediation, etc., all of which are tools that the Canton planned to disseminate to its FDs in order to facilitate management processes.

The principal cultural references were: the concept of added value for the organisation (Bisio 2002; Amietta and Amietta 1996; Quaglino 1979), setting priorities (Morgenstern 2004), organisational re-engineering (Hammer and Champy 1993; Lawler and Ledford 1993; Perrone 1990), knowledge management (Senge 1990) and the skills approach (Boyatzis 1982; McClelland 1973; Spencer and Spencer 1993; Camuffo 1997).

12.2.3 Human Resource Management Competencies

The ability to manage human resources refers to the aptitude/ability (innate or acquired) to manage cantonal employees.

In the training programme, this was divided into two components:

- Knowledge of the human resource management operating systems (Costa 1990; Solari 2004; Mabey and Salaman 1995; Noe et al. 2006), and knowledge of the Cantonal Administration internal regulations;
- Leadership, understood as the ability of the public manager to influence her/his collaborators. The main cultural references for this aspect are found in situational leadership theory (Hersey and Blanchard 1977), contingency theory (Fiedler 1967) and transformational leadership theory (Burns 1978 and Bass 1985).

12.2.4 Communication Skills

Communication skills refer to the aptitude/ability (innate or acquired) to convey concepts, information and emotions both inside and outside the organisation.

In the training programme, the concept of communication skills was divided into two components:

- Internal communication: understood as the ability to manage relationships in the workplace by means of appropriate communication and meta-communication methods (Bateson 1972; Brandler et al. 1982)
- External communication: relations with the media and communication with the “general public”. This part of the training programme was very practical (with little reference to ideas and concepts found in the literature) and was conducted at the RSI (the national Ticino television and radio station).

The aim of the training impact assessment system was to understand if, in these four areas, there was an improvement in managerial performance.

12.3 Research Questions

The present chapter is a case study of a training impact assessment model that was formulated specially for a public administration context.

The specific questions asked were:

“Did the FDs achieve the training target-competencies?”

“Did the FDs managerial behaviour match the expected managerial model?”

It should be noted that the chapter refers to an initial test that included only one questionnaire (ex-post), using a limited number of FDs (17). Therefore, in view of this statistical population, the responses processed had a relatively reduced level of reliability.

There were two reasons for conducting this test:

- To obtain an initial feed-back on the courses;
- To test the assessment tools that are planned to be used when the training programme is fully operational.

It was envisaged that the assessment system be applied extensively, with an ex-ante and ex-post analyses of the data. A full understanding of the limits and potential of the training programme would be possible only after these analyses had been completed.

12.3.1 Research Methodology

In order to evaluate the impact made by the training programme on the four managerial skills (entrepreneurship, efficiency, human resources and communications) the Kirkpatrick model (1994) was applied.

The next part of the present chapter illustrates the way in which the statistical population was identified, and the tools adopted, followed by the expression of some specific considerations regarding the system used to evaluate managerial behaviour in the workplace.

12.3.2 Statistical Population

In terms of the population investigated, it is useful to give some initial comments regarding its composition, and the validity of the observations and the relative limits.

12.3.3 Composition of the Population

The Ticino Cantonal Administration workforce (excluding magistrates) consists of 271 Public Managers (266.4 expressed as employment percentage), of whom 245 are male and 26 female. Almost all are employed on full time contracts, as shown in Fig. 12.4.

All participants of the training programme were required to certify their skills and complete the assessment questionnaire, thus providing comprehensive data on the analysed population (100% of the FD population).

Department	No. FD	Men	Women	F.T.E.
CHANCELLERY	8	7	1	7.3
DECS	38	28	10	37.9
DFE	70	64	6	69.4
DI	55	49	6	53.5
DSS	25	24	1	23.7
DT	74	72	2	73.6
CCF	1	1	0	1
Total	271	245	26	266.4

Fig. 12.4 Division of public managers employed by the Ticino Cantonal Administration. *Source:* data processed by the authors on the basis of USM (2011) data

12.3.4 Validity

Since it would have been impossible to assess the impact of evaluation process without any external influences, any changes observed, particularly those regarding behaviour in the work place, may have resulted from phenomena that could not be ascribed to the programme; nevertheless, the analysed target was equivalent to 100% of the observable population, consisting of persons operating in different contexts⁵, subjected to diverse environmental influences⁶, in order that the results may be considered as a sufficiently reliable measurement of the impact made by the training.

In effect, the training programme appeared to be the only variable common to all the FDs; therefore, any behavioural change following the course may reasonably be ascribed to the impact of the training programme.

⁵Each FD made her/his decisions in full autonomy, working under different managers and with different collaborators, so the influences affecting those who completed the questionnaires differ for each member of the statistical population.

⁶The reference environments are diversified in terms of:

1. Level of certainty/uncertainty related to:
 - Clarity of tasks
 - Difficulty of the work executed by them
2. Feed-back time for finding out the outcome of action taken;
3. Level of influence on the people occupying other organisational roles, in terms of:
 - Strategic success of the business
 - Relative importance of each environmental subsystem (Lawrence and Lorsch 1970).

In addition to this, the environments do not affect each other and are in different developmental contexts (some departments are developing strongly, while others are being redefined/resized, etc.).

12.3.5 Investigation Tools

Three different tools were used in order to observe the phenomenon (the impact made by the training programme, based on Kirkpatrick's three levels):

1. In order to understand the level of satisfaction with the training course (*reaction*), a questionnaire was designed and distributed at the end of the course (Attachment 2).
2. In order to assess the *learning*, the results of the end-course exams were examined (Attachment 3).
3. In order to understand the change in workplace *behaviour*, an ad hoc questionnaire was designed (Attachments 4–5–6).

The present chapter focuses particularly on the third point, which reveals the most interesting research elements. It is, therefore, necessary to describe the construction methodology utilised for the workplace behaviour assessment questionnaire.

12.3.6 Workplace Behaviour Assessment

The system used, in order to assess the impact made by the training programme on working behaviour, was based on three questionnaires given to three statistical populations: the person directly involved (the FD participating in the programme), her/his direct manager (SG) and her/his collaborators (COLL).

Each population was asked to assess the FDs' managerial behaviour in the workplace.

12.3.7 All-Round View

The aim was to obtain an all-round view, in the sense of involving various hierarchical levels: namely, the person directly involved, her/his direct manager and staff.

This system allowed each person, when completing the questionnaire, to express her/his "subjective evaluations" (in this case, regarding the way in which the FD interprets her/his managerial role). By comparing the various evaluations, it was possible to draw up a "relatively objective picture" of the behaviour of the various FDs.

Three questionnaires were used in order to gather the data: direct manager, collaborator(s) and person directly involved. The questionnaires followed the same logical structure, but the questions differed in accordance with the role of the respondent.

12.3.7.1 Ex-Ante and Ex-Post Analyses

An initial observation was completed before the beginning of the training programme, and a second at the end of the programme: the persons involved and questions asked in both these phases were the same. Since the time interval between the ex-ante and ex-post observations was 3 years (equivalent to the duration of the training programme), some changes in staff may have occurred during this period (some of those who completed the first questionnaire might subsequently have changed position or left their job). If this were the case, the ex-ante questionnaire was eliminated.

12.3.7.2 Control Questions

Each observed topic was investigated by means of one direct question and two control questions, each of which was answered using a scale of 1–6. If the same question was answered in a manner that was incoherent, it was classified as unreliable.

The Likert (1932) method was used to create the scale. However, in order to clearly separate the negative from the positive answers, an equal numbered interval scale was used.

12.3.8 *Structure of the Questionnaire*

The questionnaires were designed in order to assess the transfer of the four managerial skills (identified during the training needs analysis phase⁷ into behaviour implemented in the workplace.

A number of different courses were organised for each managerial skill and are listed in Fig. 12.5, together with details regarding their duration.⁸

To measure the extent to which skills are transferred into real workplace behaviour, a classification was made of workplace attitudes and behaviours in order to demonstrate that the FDs possessed the managerial skills.

With regard to the four managerial skills, 11 items were identified (skills and behaviours that an FD should have), each of which was measured (as described above) by means of three statements the respondents were asked to indicate their level of agreement/disagreement, on a scale of 1–6, where:

⁷See Chap. 2; note that these skills derive from the suggestions made by the FDs during the shared design process.

⁸N.B. some courses focus on more than one skill. For example, the “Delegating” course develops the skills associated with human resource management, together with those related to organisational efficiency. The managerial skills to which the courses refer were defined by examining the topics confronted.

<i>Managerial Skill</i>	<i>Course</i>	<i>Duration (in 8-hour days)</i>
Tendency to Public Entrepreneurship	The role of the Public Manager *	2
	Administrative Law **	1
	Understanding the centralised services **	½
Tendency to Organisational Efficiency	Basic Organisation (processes and structures)	2
	Time Management	1
	Decision Making & Problem Solving	2
	Effective Meeting Management	1
	Delegating***	1
Human Resource Management Skills	Motivation and Team Building***	2
	Selection and Hiring Processes ***	1
	Personnel Procedures and Regulations**	1
Communication Skills	Presentation and Communication Techniques ***	1
	Relations in the Workplace	2

* external course, lasting more than 8 hours.

** in addition to the classroom time specified (shown in the third column), these courses also required further practice conducted in distance learning mode.

*** in addition to the classroom time specified (shown in the third column), these courses also required specific practice, which may be completed with the assistance of a coach (if requested by the course participant).

Fig. 12.5 Course structure and managerial skills. *Source:* data processed by the authors, on the basis of USM (2016) data

- 1 and 2 represent strong disagreement;
- 3 represents moderate disagreement;
- 4 represents moderate agreement;
- 5 and 6 represent strong agreement.

Figure 12.6 summarises the general logical connections between the questions within the questionnaire, the items, and the managerial skills that were the objective of the “Lead and Manage” training programme.⁹

All the managerial skills had three assessment items, except “Communication skills, which had only two items. This was due to a specific choice made by USM, which decided to measure only those communication skills that were related to presentations in work teams and at meetings.

After processing the responses, the following analyses were conducted:

- A quantitative analysis of the answers;
- Reliability analyses (Kruskal–Wallis test, 1952);
- Standard deviation.

⁹N.B. some assessment items indicate more than one skill: for example, “delegate” can be associated with either “Human Resource Management Skills” or “Tendency to Organisational Efficiency”.

In order to obtain an unequivocal assessment of the impact made by the training courses on the skills, we decided to ascribe only one skill, based on the actual contents of the lessons.

<i>Managerial Skill</i>	<i>Item for assessment</i>	<i>Question Numbers</i>
Tendency to Public Entrepreneurship	Tendency to Change	9; 13; 30
	Tendency to the service	3; 17; 26
	Decision taking / autonomy	5; 6; 21
Tendency to Organisational Efficiency	Process Analyses	18; 25; 29
	Defining Priorities	14; 24; 33
	Delegating	1; 4; 16
Human Resource Management Skills	Motivation	2; 8; 19
	Team Building	12; 20; 28
	Selection of resources / skills	10; 23; 32
Communication Skills	Communicating contents	11; 15; 27
	Utilisation of communication tools	7; 22; 31

Fig. 12.6 Managerial skills, assessment and questions. *Source:* data processed by the authors, on the basis of USM (2016) data

12.4 Description of the Results, Critical Discussion and Potential Impact of the Research in Terms of the Development of Knowledge

To date, the initial test has been conducted on a sample of 154 questionnaires completed by respondents, the composition of which is shown below:

17 FDs (10.9%)
17 Bosses (10.9%)
122 Subordinates (78.2%)

The questionnaires referred only to the ex-post evaluation, therefore it was not possible to analyse changes in managerial styles following the course. This initial analysis acted as a test in order to verify the validity of the system. Of particular interest was the understanding of the strength of the control questions, and of an initial idea with regard to the result of the “all-round” view.

One limit of this initial analysis, however, was that the questionnaires were assessed without making any distinction in terms of the Organisational Unit (OU) of the respondents. When making the final assessment of the impact made by the training programme, the data processed should be differentiated by OU in order to obtain a more analytical understanding.

With these considerations, the initial results obtained may now be presented.

12.4.1 Data on the Level of Satisfaction

The level of satisfaction assessment was made directly by the Cantonal Administration using its own data gathering tools. Figure 12.7 shows the average

Course	Overall satisfaction
Administrative law	2.34
The role of the Public Manager	2.22
Motiving and Team Building	2.43
Delegating	2.15
Personnel Procedures and Regulations	2.15
Basic Organisation (processes and structures)	2.11
Effective Meeting Management	2.13
Presentation and Communication Techniques	2.71
Time Management	2.09
Decision Making & Problem Solving	2.17
Relations in the Workplace	2
Selection and Hiring Processes	2.13
Average result	2.22

Fig. 12.7 Level of satisfaction with the courses (6 editions). *Source:* data processed by the authors, on the basis of USM, (2015)

score for the question on “general satisfaction” with the courses. A scale of 1–4 was used, where values 0 and 1 represent two levels of dissatisfaction, while 2 and 3 represent satisfaction.

The data collected show a “positive” average level of satisfaction for all the courses (the lowest result being 2.00), and the detailed analysis of each edition shows that, overall, the participants said that they were generally dissatisfied (level of satisfaction lower than 2) with only 4 out of a total number of 48 courses (8%). This result was considered to be generally positive, considering that the courses were compulsory and that the participants did not volunteer to attend them.

12.4.2 Data on Learning

The next part of the present chapter shows data regarding the results of the course content learning tests. Figure 12.8 shows the results for the first two editions of the courses.

The results were exceptionally positive, perhaps because the first two editions were experimental and (particularly for the first edition) the participant selection process focussed on the *best performers* of the Cantonal Administration.

While the Administration acknowledged this explanation, it felt that there had been some “lenience” in terms of assessment. The teachers were therefore asked to be stricter in the future, and, in effect, the scores for the subsequent editions were different (Fig. 12.9).

P = Pass / F = Fail

Course	Edition 1		Edition 2		Pass		Fail	
	P	F	P	F	n	%	n	%
Administrative law	9	1	11	1	20	91 %	2	9 %
The role of the Public Manager	9	0	15	0	24	100 %	0	0 %
Motiving and Team Building	10	0	15	0	25	100 %	0	0 %
Delegating	10	0	13	0	23	100 %	0	0 %
Personnel Procedures and Regulations	10	0	13	0	23	100 %	0	0 %
Basic Organisation (processes and structures)	10	0	11	0	21	100 %	0	0 %
Effective Meeting Management	10	0	16	0	26	100 %	0	0 %
Presentation and Communication Techniques	10	0	12	0	22	100 %	0	0 %
Time Management	11	0	15	0	26	100 %	0	0 %
Decision Making & Problem Solving	11	0	14	0	25	100 %	0	0 %
Relations in the Workplace	9	0	13	0	22	100 %	0	0 %
Selection and Hiring Processes	12	0	17	0	29	100 %	0	0 %
Average result	121	1	106	1	184	99 %	2	1 %

Fig. 12.8 Learning in the first 2 editions. *Source:* data processed by the authors on the basis of USM (2015)

P = Pass / F = Fail

Course	Edition 3		Edition 4		Edition 5		Edition 6		Pass		Fail	
	P	F	P	F	P	F	P	F	n	%	n	%
Administrative law	14	1	14	0	14	0	8	0	50	98 %	1	2 %
The role of the Public Manager	14	0	13	0	14	0	14	1	55	98 %	1	2 %
Motiving and Team Building	12	1	12	1	12	2	8	1	44	90 %	5	10 %
Delegating	14	0	12	0	12	2	10	1	48	94 %	3	6 %
Personnel Procedures and Regulations	15	0	11	1					26	96 %	1	4 %
Basic Organisation (processes and structures)	13	2	12	3					25	83 %	5	17 %
Effective Meeting Management	11	0	10	3					21	88 %	3	13 %
Presentation and Communication Techniques	7	0	11	0					18	100 %	0	0 %
Time Management												
Decision Making & Problem Solving												
Relations in the Workplace												
Selection and Hiring Processes												
Average result	100	4	95	8	52	4	40	3	287	93 %	19	7 %

Fig. 12.9 Learning in the last 4 editions. *Source:* data processed by the authors on the basis of USM (2015)

12.4.3 Data on Behaviour

The data analyses of the questionnaires verifying the change in behaviour in the workplace was divided into three components:

- 12.4.3.1 Response coherence analyses;
- 12.4.3.2 Check of any changes in managerial behaviour;
- 12.4.3.3 Considerations regarding the training course.

12.4.3.1 Response Coherence Analyses

Response coherence analyses were conducted before proceeding to the subsequent assessments. The aim being to check:

- Internal coherence: comparison, inside each questionnaire, between the responses to the three control questions;
- External coherence: comparison between the responses of the three levels interviewed (direct manager, collaborators, course participant).

The equivalence test of the average scores in the analyses areas (Kruskal and Wallis 1952) summarises two coherences. For the categorical variables, the statistical significance of the average equivalences between groups was investigated using the chi-quadro test, while the continuous variables were subjected to the Kruskal–Wallis test, which is the non-parametric analogue of the ANOVA test. The Kruskal–Wallis test was used as it did not require a normal data distribution and could, therefore, check the equivalence of the average scores of different groups: the closer the test value was to 1, the more equal to each other would be the averages of the groups. Figure 12.10 shows the average, the standard deviation and the Kruskal–Wallis test results for the 11 items.

The table in Fig. 12.10 highlights the results in the 11 analysis areas, each consisting of three questions, where the *range* of the results per area varies from 3

		μ	σ	σ^* (VC)	KW test
Process analyses	Bosses (SG)	14.07	2.49	0.177	0.9410
	FDs (FD)	14.12	2.44	0.173	
	Subordinates (COLL)	14.07	2.56	0.182	
Specifying priorities	Bosses (SG)	13.33	3.17	0.238	0.5532
	FDs (FD)	14.70	1.68	0.114	
	Subordinates (COLL)	14.07	2.81	0.200	
Delegating	Bosses (SG)	10.62	2.09	0.197	0.2421
	FDs (FD)	11.82	3.26	0.276	
	Subordinates (COLL)	11.37	2.21	0.194	
Motivation	Bosses (SG)	12.30	2.09	0.170	0.4633
	FDs (FD)	13.29	2.02	0.152	
	Subordinates (COLL)	13.11	2.78	0.212	
Tendency towards change	Bosses (SG)	13.71	3.79	0.276	0.0876
	FDs (FD)	15.17	2.48	0.163	
	Subordinates (COLL)	13.31	3.40	0.255	
Tendency towards the service	Bosses (SG)	12.21	2.69	0.220	0.3190
	FDs (FD)	11.00	2.00	0.182	
	Subordinates (COLL)	11.42	2.02	0.177	
Decision-taking/autonomy	Bosses (SG)	12.06	2.46	0.204	0.0018
	FDs (FD)	12.70	2.31	0.182	
	Subordinates (COLL)	10.72	2.02	0.188	
Selection of resources /skills	Bosses (SG)	14.66	2.38	0.162	0.8081
	FDs (FD)	15.37	1.20	0.078	
	Subordinates (COLL)	14.93	2.63	0.176	
Team building	Bosses (SG)	13.62	2.60	0.191	0.4504
	FDs (FD)	14.70	1.75	0.119	
	Subordinates (COLL)	13.62	2.98	0.219	
Communicating contents	Bosses (SG)	14.68	2.62	0.178	0.6541
	FDs (FD)	14.52	1.77	0.122	
	Subordinates (COLL)	14.76	2.75	0.186	
Utilisation of communication tools	Bosses (SG)	13.07	3.97	0.304	0.6853
	FDs (FD)	11.93	4.12	0.345	
	Subordinates (COLL)	11.94	4.08	0.342	

Fig. 12.10 Reliability analyses of the 11 items. Source: data processed by the authors

		μ	σ	σ^* (VC)	KW test
Human resource management skills (max 54)	(SG)	40.76	6.66	0.163	0.6522
	(FD)	43.37	3.00	0.069	
	(collaborators)	42.28	6.76	0.160	
Communication skills (max 36)	(SG)	28.07	5.62	0.200	0.7580
	(FD)	26.37	5.14	0.195	
	(collaborators)	26.67	6.28	0.235	
Tendency towards organisational efficiency (max 54)	(SG)	37.36	5.20	0.139	0.2209
	(FD)	40.75	4.89	0.120	
	(collaborators)	39.65	4.79	0.121	
Tendency towards public entrepreneurship (max 54)	(SG)	38.15	7.25	0.190	0.1420
	(FD)	38.62	3.63	0.094	
	(collaborators)	35.74	5.91	0.165	

Fig. 12.11 Reliability analyses of the four skills. *Source:* Data processed by the authors

(minimum) to 18 (maximum).¹⁰ Considering the median of the results (10.5), it was immediately clear that all the groups assigned the areas with a generally high value, as confirmed by the average result of 13.24 (with an average standard deviation of 2.59).

In the variables for the “Process analyses”, “Selection of resources/skills”, “Communicating contents” and “Utilisation of communication tools”, the averages of the three groups were very similar (KW > 0.65). Moreover, in the “Process analyses” the averages were considered as almost identical (KW 0.94). “Tendency towards change” was an area that was statistically acceptable, but with test values that were more moderate. In short, the evidence emerging from the questionnaire may be considered reliable, as it was confirmed by three statistical populations.

The only item to have recorded statistically significant differences was “Decision making/autonomy” (level of significance 0.018), which, strictly interpreting the Kruskal–Wallis test (Kruskal and Wallis 1952), shows that the three groups gave responses that were not coherent. This meant that the observers had different perceptions of the FDs’ level of autonomy in terms of decision taking. Although, statistically speaking, this score gave a negative result (incoherence between the responses of the three groups), from an organisational aspect it generated interesting repercussions that could lead to deeper qualitative investigation into the reasons for the variance.

If this result were to be repeated in the next data-collection periods, it should be studied in greater detail.

At this point, the four managerial skills were analysed: “Human resource management skills”; “Communication skills”; “Tendency towards organisational efficiency” and “Tendency towards public entrepreneurship”. It was possible to evaluate only the possession¹¹ of these skills based on the perceptions of the Public Manager (FD), her/his direct manager (SG), and her/his collaborators.

Based on the existing relationship between item and skills (see Fig. 12.6), Fig. 12.11 shows the average, the standard deviation, and the Kruskal–Wallis test results.

¹⁰It should be noted that the value ascribed to each question varies from 1 to 6. The responses of the FDs, the SGs and the collaborators are grouped together.

¹¹Here, one is forced to talk about the “possession” and not “acquisition” of skills, because as this is a control sample, it is not possible to make an “ex-ante ex-post” comparison.

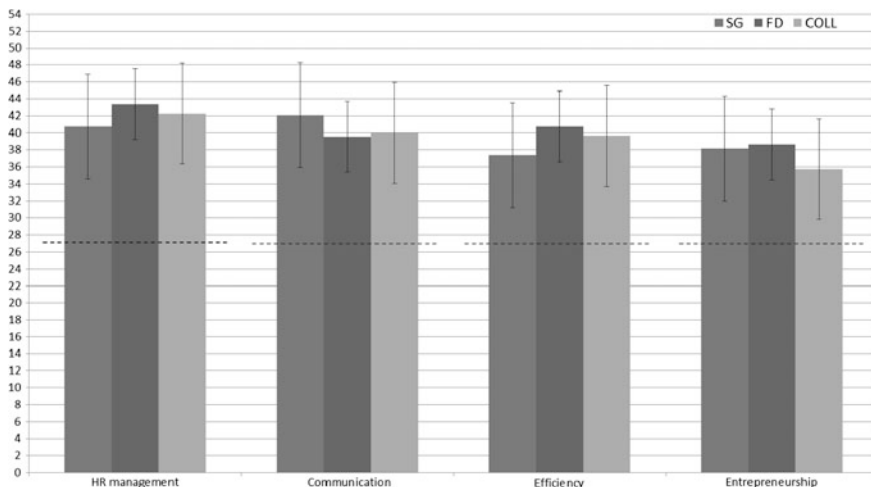


Fig. 12.12 Reliability analyses of the four skills. *Source:* data processed by the authors

SKILLS	PERCEIVED PERCEPTION
“Human resource management skills” (average absolute score of 42.1 out of 54)	77.8%
“Communications skills” (average absolute score of 27.0 out of 36)	75.0%
“Tendency towards organisational efficiency” (average absolute score of 39.3 out of 54)	72.3%
“Tendency towards public entrepreneurship” (average absolute score of 37.5 out of 54)	68.5%

Fig. 12.13 Level of possession of the skills (perception of the 2 populations). *Source:* data processed by the authors

The results obtained for the macro-skills (Attachment 9), checked with the statistical test, indicated substantial equivalence between the averages of the analysed groups ($KW > 0.05$). Therefore, it may be said that the “perception of possession of the four skills” was, on average, evaluated similarly in the three groups and that the data obtained with regard to the level of skill possessed by the FDs could be considered as credible.

This data is shown, in graph form, in Fig. 12.12.

Figure 12.12, shows clearly that the perception of the FDs’ possession of the skills (verified in the 3 groups) is, on average, good. Figure 12.13, shows the details of the results¹².

¹²This table is an extremely condensed indicator obtained by taking a simple average of the 3 averages (FD, SG and COLL) of the assessment for each skill.

The “Communication Skills” analysis was aimed at observing the ability to communicate personal opinions clearly, using some technological supports.

The high value of this score suggested that the three statistical populations recognise the possession of this skill. If this score is compared with the related learning and satisfaction results of the courses, there is confirmation that communication skills are effectively possessed.

The relatively limited FD standard deviation score (3) for “Human resource management skills” suggests that all the FDs were aware that they possessed the skills required in order to select collaborators and create a well-performing and motivated work group. This was confirmed by the SGs and the collaborators (based on the evidence of the KW test).

“Tendency towards public entrepreneurship” was a new topic introduced within the Cantonal Administration, and was the aspect that was most difficult to understand and accept. Therefore, it was not surprising that this was the skill where the Kruskal–Wallis test results were generally lower and where one assessment item (“Decision taking/autonomy”) recorded a value indicating a statistically significant difference between the participants. The reasons for this difference (in a qualitative evaluation) appeared to be connected to the “newness”, making it difficult to recognise the knowledge and skills acquired.

If the same results were to be recorded in the next data collection periods, it would be advisable to conduct further analyses.

12.4.3.2 Check of the Variances in Managerial Behaviours

Such a check of the variances in managerial behaviours would result from a comparison of the ex ante and ex-post questionnaires.

This extremely important information may not be obtained at present as the ex-ante data are lacking (they will, however, be available at the end of the next period).

12.4.3.3 Considerations Regarding the Training Programme

The final (and most important) information that may be obtained from this training evaluation model is related to the effectiveness of the “Lead and Manage” programme in terms of changing the managerial behaviour of the FDs.

Although, to date, there has been no ex-ante information ex-ante to compare with the ex-post results of the questionnaire, it was possible to assess the level reached by the FDs in terms of: Reaction, Learning and Behaviour. This information proved to be immediately useful in order to intervene on the personnel management systems (an ongoing revision of the entire system of the cantonal remuneration system based on the “Pay-for-Competencies” model).

This limitation of the evaluation, which was based on the perceptions of the respondents and not on objective data, due to the lack of an ex-ante/ex-post

analysis, should, however, be overcome at the end of the next period when ex-ante data will be available.

Another aspect that was not considered by the evaluation model adopted by the Canton was the impact of motivational variables (suggested by career and job attitudes, organisational commitment, decision/reaction to training, post-training intervention) on the process of knowledge transfer and the individual characteristics of those undergoing training: the locus of control (Baumgartel et al. 1984; Noe and Schmitt 1986; Rotter 1966) and the self-efficacy (Bandura 1986; Quinones and Ehrenstein 1997; Gist et al. 1991). However, it was possible to express some considerations with regard to the level of course satisfaction and learning.

1. Entrepreneurship orientation

The course that was most strongly related to “tendency towards entrepreneurship” was the one entitled “The role of the Public Manager”, though a qualitative examination of the level of satisfaction data produced some contrasting results.

The course that was most strongly related to “entrepreneurship orientation” was the one entitled “The role of the Public Manager”.

A further qualitative examination of the level of satisfaction data also produced some contrasting results: the end-course questionnaires show a high level of “general satisfaction” for each edition (an average of 2.22 out of 3 for the 12 courses monitored), confirmed by all the other points related to didactics and clarity of explanation. However, strongly negative evaluations were recorded for two aspects:

- Applicability of the topics to the work environment
- Level of correspondence between the contents and professional interests

With regard to the question of “applicability to the work environment” 5 editions out of 6 expressed perplexity, and regarding the “level of correspondence between the training and professional interests”, half (3 out of 6) of the participants stated that there was little correspondence.

This aspect highlighted the difficulty of trying to convey innovative concepts to a public management that is firmly consolidated on professional-bureaucratic type values, strongly suggesting that some changes be made to both the training course and to the general Cantonal organisational model.

2. Need to improve the learning evaluation systems

Learning evaluation in adult training programmes is always a sensitive topic (Knowles et al. 2008), however, after the first two editions, some corrective action was considered necessary. The monitoring system had a number of weak points: in fact, it did not monitor any corrections made by the FDs to their tests when they were given the opportunity to review their examinations and improve their results.

Further investigation, therefore, is necessary in order to gain a deeper understanding of the assessment dynamics.

12.5 Application Repercussions for Managers and Policy Makers

Once the evaluation process is operating fully (ex-ante and ex-post data), it will be possible to draw up analytical considerations regarding the training programme, checking what impact it has had on FD management styles, and, if necessary, what changes to make.

The possible applications for the Cantonal Administration are clear, both in terms of revisions to the training programme and any other organisational changes aimed at disseminating critical managerial skills to FDs.

There are two reasons why the authors of the present chapter believe the case of Canton Ticino to be particularly significant and of merit:

1. It is an example of how training may be utilised as an activation lever for change (Rebora and Meneguzzo 1990; Martone 2007; Rebora and Minelli 2007) and how it may support the organisational transition (Kotter and Schlesinger 2008);
2. The important role assigned to training in the modernisation process of Public Administration. In this sense, the case of Canton Ticino is unique within the Swiss Confederation and is infrequently found outside Switzerland. While many Swiss Cantons organise compulsory courses for their managers, and many Cantons specify end-course learning assessments, Canton Ticino is the only one to have implemented both these features. This decision has generated extraordinary consequences: any FD who, after the specified two re-sit examinations, fails to pass the test is demoted from her/his managerial position. Canton Ticino has put itself forward as the most advanced experimentation site within the Swiss Confederation.

The authors also believe that the training programme could be replicated outside of the Ticino context: the managerial skills identified (and appropriately modified to individual requirements) may be generalised for most public administration bodies (in continental European systems, at least). The differences between the various legal systems should not make any dramatic impact on the general applicability of the model and the structure of the evaluation process and the tools adopted for the purpose of acquiring information may be re-utilised in any context.

References

- Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: Thirty years later. *Personnel Psychology*, 42(2), 331–342.
- Alliger, G. M., Tannenbaum, S. I., Bennett, W., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology*, 50, 341–358.
- Amietta, P. L., & Amietta, F. (1996). *Valutare la formazione*. Milano: Unicopli.
- Ansoff, H. I., & Brandenburg, R. G. (1971). A language for organization design. *Management Science*, 17(12), B-705.

- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall, NJ: Englewood Cliffs.
- Bass, B. M. (1985). *Leadership and performance*. New York: Free Press.
- Bateson, G. (1972). *Step to an ecology of mind*. Chicago: University of Chicago Press.
- Baumgartel, H., Reynolds, M., & Pathan, R. (1984). How personality and organizational climate variables moderate the effectiveness of management development programmes: a review and some recent research findings. *Management and Labour Studies*, 9, 1–16.
- Bisio, C. (2002). *Valutare in Formazione. Azioni, significati e valori*. Milano: Angeli.
- Boyatzis, R. (1982). *The competent manager: A model for effective performance*. New York: Wiley.
- Brandler, R., Grinder, G., & Dilts, R. B. (1982). *Programmazione neurolinguistica. Lo studio della struttura dell'esperienza soggettiva*. Roma: Astrolabio.
- Burns, J. M. (1978). *Leadership*. New York: Harper and Row.
- Bushnell, D. S. (1990). Input, process, output: A model for evaluating training. *Training and Development Journal*, 44(3), 41–44.
- Camuffo, A. (1997). La Natura delle Competenze Manageriali. *Economia and Management*, 3, 31–49.
- Cannon-Bowers, J. A., Salas, E., Tannenbaum, S. I., & Mathieu, J. E. (1995). Toward theoretically based principles of training effectiveness: a model and initial empirical investigation. *Military Psychology*, 7, 141–164.
- Canton Ticino website. (2016a). Il Gran Consiglio, Retrieved January 25, 2016 from: <http://www4.ti.ch/poteri/gc/parlamento/presentazione/>.
- Canton Ticino website. (2016b). Il Consiglio di Stato. Retrieved January 25, 2016 from <http://www4.ti.ch/poteri/cds/attivita/>.
- Canton Ticino website. (2016c). Cancelleria dello Stato. Retrieved January 25, 2016 from <http://www4.ti.ch/can/chi-siamo/presentazione/>.
- Canton Ticino website. (2016d). Controllo Cantonale delle finanze. Retrieved January 25, 2016 from <http://www4.ti.ch/di/ccf/ufficio/>.
- Costa, G. (1990). *Economia e direzione delle risorse umane*. Torino: UTET.
- Consiglio di Stato. (2015). Rendiconto del Consiglio di Stato 2015, Allegato statistico. Retrieved February 18, 2016 from http://www4.ti.ch/fileadmin/CAN/TEMI/RENDICONTOCDS/2015/RENDICONTO/Allegato_statistico_2015.pdf.
- Constitution of the Republic and Canton of Ticino (Costituzione della Repubblica e Cantone Ticino) of 14 December 1997. RS 131.229. Status as of 11 March 2015.
- Daft, R. L. (2015). *Management* (12th ed.), Boston: Cengage Learning.
- European Union (2006). Recommendation of the European Parliament and of the Council of 18 December 2006 on key competences for lifelong learning L394/10, Retrieved February 12, 2016 from: <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32006H0962&from=EN>.
- Federal Constitution of the Swiss Confederation of 18 April 1999. RS 101. Status as of 1 January 2016.
- Fiedler, F. E. (1967). *A theory of leadership effectiveness*. New York: McGraw-Hill.
- Fitz-Enz, J. (1994). *How to measure human resource management*. New York: McGraw-Hill.
- Ford, J. K., & Kraiger, K. (1995). The application of cognitive constructs and principles to the instructional systems design model of training: Implications for needs assessment, design, and transfer. *International Review of Industrial and Organizational Psychology*, 10, 1–48.
- Gist, M. E., Stevens, C. K., & Bavetta, A. G. (1991). Effects of self-efficacy and post-training intervention on the acquisition and maintenance of complex interpersonal skills. *Personnel Psychology*, 44, 837–861.
- Hammer, M., & Champy, J. A. (1993). *Reengineering the corporation: A manifesto for business revolution*. New York: Harper Business.
- Hersey, P., & Blanchard, K. H. (1977). *Management of organizational behavior: Utilizing human resources* (3rd ed.). New Jersey: Prentice Hall.

- Holton, E. F. (1998). Customer service is not the best model for HRD practice. *Human Resource Development Quarterly*, 9(3), 207–209.
- Kaufman, R., Keller, J., & Watkins, R. (1995). What works and what doesn't: Evaluation beyond Kirkpatrick. *Human Resource Management Quarterly*, 5.
- Kirkpatrick, D. L., & Kirkpatrick, J. D. (1994). *Evaluating training programs*. Oakland: Berrett-Koehler.
- Knowles, M., Holton, E. F., & Swanson, R. A. (2008). *Quando l'adulto impara, Andragogia e sviluppo della persona*. Milano: Angeli.
- Kotter, J. P., & Schlesinger, L. A. (2008). Choosing strategies for change. *Harvard Business Review*, 86(7/8), 1–11.
- Kruskal, W. H., & Wallis, A. W. (1952). Use of ranks in one-criterion analysis of variance. *Journal of the American Statistical Association*, 47(260), 583–621.
- Lawler, E. E., & Ledford, G. E. (1993). A skill based approach to Human Resources Management. *European Management Journal*, 10(4), 383–391.
- Lawrence, P. R., & Lorsch, J. W. (Eds.). (1970). *Organization structure and design*. London: Irwin-Dorsey.
- Legge sull'ordinamento degli impiegati dello Stato e dei docenti (Lord) of 15 March 1995. RS 2.5.4.1. Status as of 15 March 1995.
- Likert, R. (1932). Technique for the measure of attitudes. *Archives of Psychology*, 22 (140).
- Mabey, C., & Salaman, G. (1995). *Strategic human resource management*. Oxford: Blackwell.
- Martone, A. (Ed.). (2007). *Change management: teorie organizzative e modello applicativo*. Milano: Wolters & Kluwert.
- Mathieu, J. E., Martineau, J. W., & Tannenbaum, S. I. (1993). Individual and situational influences on the development of self-efficacy: Implications for training effectiveness. *Personnel Psychology*, 46, 125–147.
- McClelland, D. C. (1961). *The achieving society*. Princeton: Nostrand.
- McClelland, D. C. (1973). Testing for competence rather than for intelligence. *American Psychologist*, 28(1), 1–4.
- Meynhardt, T. (2009). Public value inside: What is public value creation? *International Journal of Public Administration*, 32(3–4), 192–219.
- Moore, M. (1995). *Creating public value—strategic management in government*. Cambridge: Harvard University Press.
- Morgenstern, J. (2004). *Time management from the inside out: The foolproof system for taking control of your schedule—and your life*. New York: Holt.
- Noe, R. A. (1986). Trainees' attributes and attitudes: Neglected influences on training effectiveness. *Academy of Management Review*, 11(4), 736–749.
- Noe, R. A., Hollenbeck, J. R., Gerhart, B., & Wright, P. M. (2006). *Gestione delle risorse umane*. Milano: Apogeo.
- Noe, R. A., & Schmitt, N. (1986). The influence of trainee Attitudes on training effectiveness: Test of a model. *Personnel Psychology*, 39, 497–523.
- Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. Oxford, NY: Oxford University Press.
- Norman, R. (1992). *La gestione strategica dei servizi*. Milano: Etas.
- Olsen, J. H. (1998). The evaluation and enhancement of training transfer. *International Journal of Training and Development*, 2(1), 61–75.
- Perrone, V. (1990). *Le strutture organizzative d'impresa, criteri e modelli di progettazione*. Milano: EGEA.
- Polanyi, M. (1962). Tacit Knowing: Its bearing on some problems of philosophy. *Reviews of Modern Physics*, 34(4), 601.
- Prahalad, C. K., & Hamel, G. (1990). The core competence of the corporation. *Harvard Business Review*, 68(3).
- Quaglino, G. P. (1979). *La valutazione dei risultati della formazione*. Milano: Angeli.
- Quinones, M. A., & Ehrenstein, A. (1997). *Training for a rapidly changing workplace: Applications of psychological research*. Washington, DC: American Psychology Association.

- Rebora, G., & Meneguzzo, M. (1990). *Strategia delle amministrazioni pubbliche*. Torino: Utet.
- Rebora, G., & Minelli, E. (2007). *Change management. Come vincere la sfida del cambiamento in azienda*. Milano: Etas.
- Rotter, J. B. (1966). Generalized expectancies for internal vs external control of reinforcement. *Psychological Monographs*, 80, 1–609.
- Rullani, E. (2004). *Economia della conoscenza. Creatività e valore nel capitalismo delle reti*. Roma: Carocci.
- Salas, E., & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual Review of Psychology*, 52, 471–499.
- Senge, P. M. (1990). *The fifth discipline: The art and practice of the learning organization*. New York: Currency Doubleday.
- Solari, L. (2004). *La gestione delle risorse umane - Dalle teorie alle persone*. Roma: Carocci.
- Spencer, L. M., & Spencer, S. M. (1993). *Competence at work: Models for superior performance*. New York: Wiley.
- Tannenbaum, S. I., Mathieu, J. E., Salas, E., & Cannon-Bowers, J. A. (1991). Meeting trainees' expectations: The influence of training fulfillment on the development of commitment, self-efficacy, and motivation. *Journal of Applied Psychology*, 76(6), 759–769.
- Van Buren, M. E., & Erskine, W. (2002). *The 2002 ASTD state of the industry report*. Alexandria, VA: American Society of Training and Development.
- Warr, P., Bird, M., & Reckham, N. (1970). *Evaluation of Management Training*. London: Gower Press.
- Winfred, A., & Winston, B. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features.
- Worthen, B. R., & Sanders, J. R. (1987). *Educational evaluation*. New York: Longman.

Chapter 13

Analysing Corruption: Effects on the Transparency of Public Administrations

Isabella Fadda, Paola Paglietti, Elisabetta Reginato and Aldo Pavan

Abstract There is a vast body of literature supporting the claim that the availability and accessibility of information play a vital role in contrasting corruption. Bastida and Benito (2007) demonstrate that the less corrupt a country is, the higher its level of budget transparency is, entailing that is not transparency which curbs corruption but rather the other way around. In this latter perspective the present study tries to contribute to the debate about transparency and corruption through the analysis of a case related to the diffusion of corruption in the twenty Italian regions. The study demonstrates that in regions with higher levels of corruption public administrations commitment towards transparency is lower compared to regions with inferior corruption levels.

Keywords Corruption · Transparency · Public administrations · Italy

13.1 Introduction

Tanzi argues that if *corruption could be measured, it could probably be eliminated* (Tanzi 1998, p. 576), however given the relevance of this phenomenon many have tried to quantify it—academics, NGOs, private enterprises. Most of the measures developed are survey-based indicators representing proxies for the spread of cor-

I. Fadda (✉) · P. Paglietti · E. Reginato · A. Pavan
Dipartimento di Scienze Economiche e Aziendali, Università degli Studi di Cagliari,
Viale S. Ignazio 17, 09123 Cagliari, Italy
e-mail: isafadda@unica.it

P. Paglietti
e-mail: ppagliet@unica.it

E. Reginato
e-mail: reginato@unica.it

A. Pavan
e-mail: apavan@unica.it

ruption at the national level—such as the Corruption Perceptions Index (CPI) developed by Transparency International—which in their earlier version only accounted for corruption perceptions, while current ones also assess the actual experience of respondents with acts of corruption. Criticisms have been raised regarding the reliability and adequacy of these indicators (see, e.g.: Golden and Picci 2005; Knack 2006; Kurtz and Schrank 2007) for a number of different reasons. Yet, the development of these corruption measures has allowed the production of a vast and ever-growing body of literature investigating its causes and consequences as well as remedies against it.

It is a well established fact that corruption finds a fertile ground in the lack of transparency; as Stiglitz (2002b, p. 35) claims *secrecy is the bedrock of persistent corruption...sunshine is the strongest antiseptic*. Consistently, the relation between corruption and transparency has long been investigated, however most of the existing literature has been produced following the release of the CPI or other analogous aggregate measures of perceived corruption and thus it is subject to the same aforementioned criticisms.

Among the criticisms, Seligson (2006) emphasises that national aggregate measures of corruption assume that corruption is uniformly spread across each single nation, thus overlooking the fact that for different reasons—socioeconomic, demographic, etc.—specific geographic areas exist within a country wherein corruption is more pervasive. As a result, in these areas, the implementation of anti-corruption programs may be hindered.

Given the relevance, from a policy making perspective, of considering the different spread of corruption within the different countries, the present research analyses the relation between corruption and transparency, in a “within-country” perspective. To this end the case of the Italian public administrations is investigated, but contrary to most of the aforementioned corruption literature, the case analysis is conducted using a non-survey based indicator of corruption. Consistently with Bastida and Benito (2007) who demonstrate that the less corrupt a country is, the higher its level of budget transparency is, it is assumed that regions affected by higher levels of corruption are also characterised by a lesser commitment toward transparency and hence by lower levels of compliance to the transparency provisions.

As of 2012 the Italian legislator has put much effort in the fight against corruption, aligning the Italian legal system with the International Treaties of which it is a signee, for instance the United Nations Convention Against Corruption (UNCAC) and the Council of Europe Criminal Law Convention on Corruption. Despite these efforts, the corruption level in the country is still high and significantly above the average level of western developed countries whatever ranking is used to assess it, and this makes the Italian case particularly interesting.

Among the measures adopted in this remarkable endeavour, the research focuses on the implementation of the transparency law—which will be examined in depth later—on account of the importance of transparency as a means to curb corruption (IMF 2015; OECD 2014; United Nations 2003). Consistently with Islam (2002) the

study considers the implementation of the transparency law as indicator of the commitment of public administrations toward transparency.

The chapter discussion will be organised as follows: the analysis of the relevant literature on corruption and transparency will be reported first, after that, the research method and inquiry will be explained. Data analysis will follow and finally the discussion and conclusions will be presented.

13.2 Literature Review

Although corruption is not difficult to recognise when it is observed, the different forms it can take on complicate the task of those who try to define it. According to Transparency International corruption is: “*the abuse of entrusted power for private gain*”. Similar definitions can be found in the literature where this phenomenon is described in terms of use/abuse/misuse of public office/powers for private gain (among others: Campos and Pradhan 2007; Golden and Picci 2005; Kaufmann 2002; Kolstad and Wiig 2009; Lambsdorff 2005; Lindstedt and Naurin 2010; Tanzi 1998; Treisman 2007). The beneficiary of the private gain may as well be a single person or a group of persons such as families, friends or political parties (Tanzi 1998; Treisman 2007). Bribery, embezzlement, patronage are just some of the forms that corruption can assume ranging from petty—or bureaucratic—to grand—or political—corruption.

It is demonstrated that corruption has distortive effects on economy like poor economic growth (Mauro 1995), twisted government spending (Gupta et al. 2001; Mauro 1998) and income inequality (Gupta et al. 2002; Gyimah-Brempong 2002). Besides, non-economic consequences have been identified, which are particularly severe for poor developing countries, e.g. higher infant and child mortality rates (Gupta et al. 2001) and environmental degradation (Plummer and Cross 2006). More generally, as recently highlighted in the OECD’s Report “Boosting integrity, Fighting corruption”, corruption has implications whose costs in terms of human sufferings go beyond the mere monetary losses (OECD 2014, p. 2).

The causes of corruption are as many as the forms the phenomenon can assume and moreover it is not always easy to disentangle causes and consequences—hence the related variables—which force affected countries in a vicious cycle (Lambsdorff 2005).

Tanzi (1998), for instance, discriminates between direct and indirect causes; among the former are included discretion allowed to public officials over important decisions and problems in financing of political parties, while in the latter are reported the quality of bureaucracy and the absence of adequate institutional controls. Treisman (2007) in his review of the studies on this subject finds that the strongest evidence resulting from the literature on the causes of corruption is the one highlighting the correlation between high level of economic development—in terms of per capita GDP—and low perceptions of corruption. He also finds robust evidence that countries with a long history of liberal democracy are perceived as

less corrupt. In this last respect Montinola and Jackman (2002) find that not only political competition affects the level of corruption, but that a threshold exists which explains why dictatorships are slightly less corrupt than partially or newly democratized countries.

Most of the studies that analyse the causes and consequences of corruption are built on survey-based indexes of corruption (Andersen 2009; Bauhr and Grimes 2014; Gupta et al. 2002; Gyimah-Brempong 2002; Lindstedt and Naurin 2010; Lio et al. 2011; Mauro 1995; Montinola and Jackman 2002) and not on objective measures of actual corruption. The best-known of these indexes is probably Transparency International's Corruption Perceptions Index (CPI) but other relevant indicators have to be mentioned here: the Economist Intelligence Unit's Business International index (BI), the World Economic Forum Global Competitiveness Index (GCI), the World Bank's Control of Corruption Index (CCI). While earlier versions of these indexes only measured corruption perceptions, current indicators also assess actual experience of respondents in relation to acts of corruption. Criticism has been raised regarding the reliability and adequacy of these indicators (see, e.g.: Golden and Picci 2005; Knack 2006; Kurtz and Schrank 2007) for a number of different reasons: over time comparability due to changes in the sources used for the construction of the indexes over the years as well as in the composition of the sample of countries analysed (Arndt and Oman 2006; Golden and Picci 2005; Knack 2006; Kurtz and Schrank 2007); sample selection problems which do not allow to capture the opinions of businesspeople who are deterred from entering local markets expressly because of corruption (Kurtz and Schrank 2007); the heavy reliance on opinions of businesspeople which allows a fair evaluation of corruption in business transactions, but not of that in the many activities pursued by private citizens (Seligson 2006); biased opinions due to cultural factors and preconceptions about corruption, and most of all opinions reflecting personal perceptions of corruption but not its actual occurrence in each country (Kurtz and Schrank 2007; Seligson 2006). In this last respect, Treisman (2007) in his review observes what he defines a *puzzling dichotomy*. As a matter of fact he finds that while perception based indicators are highly correlated with several factors commonly believed to cause corruption, these same indexes can be hardly correlated with corruption actual occurrences as measured by experience based indicators. He concludes hypothesising that this dichotomy might be due to the fact that subjective indexes do not actually measure corruption frequency but rather inferences made on the basis of conventional understandings of corruption's causes (Treisman 2007, p. 213).

Transparency is regarded as a value to which public-policies have to be inspired in order to curb corruption by major supernational organisations and nongovernmental organisations (NGOS) (see for example: IMF 2015; OECD 2014; United Nations 2003). It is among these organisations that the term transparency initially gained momentum, in particular following the establishment of "Transparency International", when its use became common among the large public as well as the academics (Ball 2009). Since then the meaning assigned to the term transparency has evolved and subsequently not univocally used in the literature, as different

authors have recently claimed (e.g. Ball 2009; Bellver and Kaufmann 2005; Harrison and Sayogo 2014; Michener and Bersch 2013; Williams 2015). An established notion of transparency is the one relating it to the availability of information (Curtin and Meijer 2006; Gerring and Thacker 2004; Grimmekhuijsen et al. 2013; Welch et al. 2005) that, on the one hand, deprives transparency of its value laden nature, on the other, sometimes leads to use the term transparency as a synonymous of openness (Meijer et al. 2012; Pallot 2001).

Some authors contend that transparency can take on different forms (Heald 2006; Meijer 2013; Reynaers and Grimmekhuijsen 2015). Heald (2006) in particular identifies what he calls *varieties* of transparency starting from the assumption that transparency can assume different forms depending on its direction—upward, downward, outward and inward. In his argument these forms of transparency can be usefully observed considering the dichotomies between event transparency versus process transparency; transparency in retrospect versus transparency in real time; nominal versus effective transparency (Heald 2006, pp. 29–35).

Williams (2015) argues that although the concept of transparency is used in the literature in a variety of ways, it is still possible to identify two recurring elements: the provision of information—which is relevant, timely, reliable, complete and understandable for its recipients—and the strengthening of public officials accountability towards relevant stakeholders (Williams 2015, p. 805).

Otenyo and Lind (2004) claim that Internet has changed our understanding of transparency and that its usage in government reinforces citizens' empowerment. Many initiatives aimed at promoting transparency and fighting corruption are being developed worldwide basing on the use of Information and Communication Technologies (ICT) (Bhatnagar 2003; Shim and Eom 2008) such as the widely studied Seoul municipality's Online Procedures Enhancement for civil applications (OPEN). Rely and Sabharwal (2009) highlight that in many countries the implementation of transparency laws is often tied to the implementation of e-government initiatives. This is the case also in Italy where, in recent years, a transparency law was issued which heavily relies on the disclosure of public administrations data on the web.

The present study tries to contribute to the debate about the link between corruption and transparency analysing the case of the twenty Italian regions.

13.3 Research Method and Objective

As previously mentioned, the purpose of the research is to investigate the relationship between corruption and transparency in a “within-country” perspective to assess whether in regions characterised by higher levels of corruption it is possible to observe a lesser commitment toward transparency. To this end the case of the Italian public administrations is analysed considering their territorial distribution among the twenty regions into which the country is divided (NUTS-2 classification).

The relevance of the Italian case arises from the significance of the problem in this nation, in absolute terms, as well as in comparison with other western developed countries. As a matter of fact, according to the Eurobarometer survey on corruption conducted in 2013, the 97% of the Italian respondents (EU average 76%) consider corruption a widespread phenomenon (European Commission Directorate-General for Home Affairs 2013, p. 6). The 2014 European Commission Anticorruption Report, as well, highlights that in spite of the efforts made in recent years to contrast corruption, this phenomenon still remains a serious challenge in Italy (Commissione Europea 2014). Furthermore in the last Transparency International's survey (Transparency International 2015) Italy obtained a score equal to 44 out of 100, which is the second worst result among EU countries (only Bulgaria scored less than Italy) wherein the average value of the CPI index is 65.

To fulfil the research objective corruption is measured using the corruption index (CI) developed by Nifo and Vecchione (2014), who defined this measure as an element of a composite indicator of the institutional quality of Italian public administrations. In the construction of this CI three elements are considered¹: the regional number of crimes committed against the public administration over the number of public servants, the regional number of local administrations overruled by the State authorities because of administrative mishandling or mafia infiltrations, on the total number of municipalities of the region, and the Golden and Picci (2005) index of corruption (Nifo and Vecchione 2014, pp. 1633–1636). The latter consists of a ratio between the amounts of physically existing public infrastructure and the amounts of money cumulatively allocated by government to create these public works (Golden and Picci 2005, p. 37). Comparing these two measures of public infrastructure allows to observe how much the government paid for the existing infrastructure in each region and to set a national average price for public constructions. The logic behind this indicator is that if the government paid for physical infrastructure more than the national average, then this indicates a waste of resources, mismanagement and fraud in the public contracting process (Golden and Picci 2005, p. 39). Although the index does not capture the extent of corruption at a single point in time, it does provide a proxy for the historically accumulated corruption in public works contracting in the years preceding the late 90s.

The resulting CI is a measure which varies in a [0,1] range where lower values of the index indicate higher values of corruption; data availability covers the years from 2004 to 2012² and the related sources are reported in Table 13.1. Besides being a proxy for actual corruption, contrary to survey based indices which assess perceived corruption, Nifo and Vecchione's CI presents another important advantage compared to aggregate measures such as the CPI index, as it provides information about the spread of corruption at the regional level thus allowing to fill the

¹For full details about the construction of the index see Nifo and Vecchione (2014).

²The full data-set is available at <https://sites.google.com/site/institutionalqualityindex/dataset>.

Table 13.1 Elements of the CI

Index element	Source	Years
Crimes against PA	ISTAT: “Indicatori territoriali per le politiche di sviluppo”	2004–2011
Overruled municipalities	Interior Ministry: “Relazione sull’attività svolta dalla gestione straordinaria dei Comuni commissariati”	1995–2012
Golden and Picci’s index	Golden and Picci (2005)	1997

Source Nifo and Vecchione (2015a)

gap highlighted by Seligson (2006) about considering corruption uniformly distributed across national territories.

For each region the respective CI is related to a transparency measure which is obtained by computing data on the compliance of Italian public administrations’ websites to the national transparency rules and standards set by the decree n. 33/2013—so called Transparency decree. This data is retrieved from the Italian web portal “The Compass of transparency”.³ Both the decree n. 33 and “The Compass of transparency” are among those initiatives implemented to foster citizenry control over the use of public resources and the consistency of Italian public administrations activities to their statutory mission. The Transparency decree was issued by the Italian parliament and imposes to all public administrations the disclosure of a large set of information concerning their organisation and management in a dedicated area of their websites that has to be labelled “*Amministrazione Trasparente*”—transparent administration. This section has to be organised into 22 first level subsections, which in turn have to be organised into another variable number of second level subsections.

The Compass of Transparency is instead an initiative launched by the Italian government, and it consists of a web portal which allows the real-time assessment of the compliance of public administrations websites to the prescriptions of the aforesaid decree. In particular this portal contains a section labelled “*colora la trasparenza*”, which means give a colour to transparency, where it is possible to assess the compliance level to the “transparency decree” of public administrations websites. The algorithm used by the portal verifies whether in the analysed websites the “transparent administration” section is present and if its structure matches the standard legal requirement for both first and second level subsections. The portal allows on demand evaluation of single public administrations as well as the assessment of groups of public administrations. In particular it is possible to obtain regional level or province level aggregated data which is then clustered by type of administration—e.g.: municipalities, schools, universities, hospitals, agencies and so on.

Consistently with Islam (2002) the study considers the implementation of the transparency law an indicator of the commitment of public administrations toward transparency, thus the presence of the “transparent administration” section in the

³See www.magellanopa.it.

public administrations' websites is considered here as a proxy for the commitment to transparency.

For each one of the 20 Italian regions the transparency level is obtained dividing the number of compliant public administrations by the number of total public administrations monitored in the region so that the index ranges between 0 (least transparent) and 1 (most transparent). At the time of data extraction the number of websites monitored was 10.967 which is more than half the total population of Italian public administrations.

13.4 Data Analysis

As the research aims to investigate whether regional differences in the distribution of corruption are related to the commitment to transparency, data was analysed at first in order to verify whether the aforementioned differences actually exist. The results of the analysis confirmed, for all the years considered, the existence of a significant difference in the spread of corruption between the northern and central regions of the country and the southern ones.⁴ More specifically in the northern and central regions it is possible to observe higher values of the CI thus indicating lower levels of corruption (Table 13.2).

As for transparency, data observation shows that the index ranges between 0.7336 and 0.9153 and that eleven out of twenty regions obtain a score that is higher than the national average (i.e. 0.8413). Eight of these eleven regions are located in the north area of the country, two are in the central area, while the remaining one is one of the two main islands. Hence, coherently with the above observed results about corruption, regional differences emerge between northern and central regions—where higher values can be observed—and southern ones. The comparison of the mean values obtained by each sub-group of regions—north, centre and south—(Table 13.3) shows significant differences between northern regions and southern ones.

Based on these results, suggesting a negative relation between transparency and corruption, the correlation analysis was at first performed considering all twenty the regions and considering the CI data for the year 2012 which is the latest available.

Because lower values of the CI indicate higher values of corruption, the transformed variable ($1 - CI$) was computed so as to obtain higher values of the index for higher levels of corruption. As it can be observed—Table 13.3—the resulting value for Pearson's r is as par as -0.44 . Data observation in the scatter plot (Fig. 13.1) however showed that Umbria presented a transparency index unexpectedly low

⁴According to the ISTAT classification: the north area includes Emilia Romagna, Friuli Venezia Giulia, Liguria, Lombardia, Piemonte, Trentino Alto Adige Valle d'Aosta, Veneto; the central area includes Lazio, Marche, Toscana and Umbria; south and islands area includes Abruzzo, Basilicata, Calabria, Campania, Molise, Puglia, Sardegna, Sicilia.

Table 13.2 Corruption—Wilcoxon rank sum test and Welch two sample *t* test

Area1	Area2	M_1	M_2	W	p value		p value
2004							
North	South	0.857	0.541	61	0.001	4.383	0.001
North	Centre	0.857	0.854	16	1	0.057	0.956
Centre	South	0.854	0.541	30	0.016	4.122	0.002
North + Centre	South	0.856	0.541	91	0.000	4.531	0.002
2005							
North	South	0.867	0.696	49	0.083	2.431	0.036
North	Centre	0.867	0.909	11	0.461	-0.851	0.423
Centre	South	0.909	0.696	29	0.028	2.863	0.017
North + Centre	South	0.883	0.696	78	0.020	2.705	0.025
2006							
North	South	0.836	0.679	51	0.04988	2.313	0.042
North	Centre	0.836	0.903	7	0.154	-1.443	0.187
Centre	South	0.903	0.679	31	0.008	3.249	0.009
North + Centre	South	0.858	0.679	82	0.007	2.756	0.021
2007							
North	South	0.860	0.678	50	0.065	2.362	0.044
North	Centre	0.860	0.920	6	0.109	-1.675	0.136
Centre	South	0.920	0.678	31	0.008	3.090	0.014
North + Centre	South	0.880	0.678	81	0.009	2.661	0.029
2008							
North	South	0.928	0.749	53	0.028	2.078	0.071
North	Centre	0.928	0.957	13	0.683	-1.155	0.275
Centre	South	0.957	0.749	32	0.004	2.471	0.041
North + Centre	South	0.938	0.749	85	0.003	2.229	0.059
2009							
North	South	0.892	0.708	54	0.021	2.269	0.054
North	Centre	0.892	0.948	5	0.073	-2.136	0.062
Centre	South	0.948	0.708	32	0.004	2.964	0.019
North + Centre	South	0.911	0.708	86	0.002	2.520	0.037
2010							
North	South	0.901	0.742	52	0.038	1.908	0.094
North	Centre	0.901	0.942	9	0.282	-1.573	0.155
Centre	South	0.942	0.742	30	0.016	2.394	0.045
North + Centre	South	0.915	0.742	82	0.007	2.091	0.073
2011							
North	South	0.895	0.703	58	0.005	2.418	0.041
North	Centre	0.895	0.927	12	0.570	-0.991	0.351
Centre	South	0.927	0.703	31	0.008	2.805	0.022
North + Centre	South	0.906	0.703	89	0.001	2.596	0.033

(continued)

Table 13.2 (continued)

Area1	Area2	M_1	M_2	W	p value		p value
2012							
North	South	0.888	0.713	53	0.029	2.196	0.06
North	Centre	0.888	0.935	10	0.368	-1.676	0.131
Centre	South	0.935	0.713	31	0.008	2.789	0.024
North + Centre	South	0.904	0.713	84	0.004	2.422	0.043
2016 estimated values							
North	South	0.886	0.795	62	0.002	5.693	0.000
North	Centre	0.886	0.840	21	0.441	1.352	0.262
Centre	South	0.840	0.795	23	0.269	1.239	0.281
North + Centre	South	0.871	0.795	85	0.005	3.937	0.001

Source own elaboration on data from Nifo and Vecchione available at <https://sites.google.com/site/institutionalqualityindex/dataset> for the years (2004–2012). Estimated data (year 2016) own elaboration

considering the respective corruption index; in view of that an outlier test⁵ was performed to assess whether Umbria's data significantly influenced the correlation analysis. The test confirmed that Umbria is an outlier (p value = 0.024) thus the correlation was computed again excluding this region and as a result Pearson's r increased ($r = -0.57$). Looking at Fig. 13.1, it is evident that the transparency index for Umbria is rather low notwithstanding a level of corruption that is almost zero. In our opinion, this result suggests that, in this case, the (low) level of transparency depends from other unobserved factors and not from corruption.

Because, as already mentioned, the latest CI data observations refer to 2012, a time gap is present in the analysis between corruption data and transparency ones. To overcome this limitation and confirm the validity of the previous findings, data about corruption in 2016 was estimated by means of an exponential smoothing analysis from the available data. For each individual region, the time series of the CI observed from 2004 to 2012 was considered and the estimated value of the CI for 2016 was obtained from a 4-steps ahead forecast performed through the exponential smoothing method. In this way, it was possible to consider the levels of estimated Corruption in 2016 and analyse the relationship between this new estimated variable and the observed levels of Transparency for 2016. Pearson's r was then computed again. Consistently with the first analysis the Umbria region appeared as an outlier, hence the outlier test was performed once again and, on this basis, the correlation was computed once more excluding this region. The results of the tests, reported in Table 13.3, confirm the strong negative correlation between corruption and transparency, thus allowing to corroborate the research hypothesis that commitment to transparency is lower where corruption is higher.

Finally, even though the proxy used to measure transparency is only a measure of formal compliance to a legal provision, which is not able to capture the quantity

⁵The test applied is the Bonferroni test for the studentized residuals obtained from a linear model fit. See among others (Weisberg 2014).

Table 13.3 Transparency index—TI

Region	Number of PAs	Total PAs	Compliance ratio (%) TI
Abruzzo	322	392	82.14
Basilicata	151	188	80.32
Calabria	179	244	73.36
Campania	650	831	78.22
EmiliaRomagna	527	592	89.02
FriuliVenezia Giulia	212	233	90.99
Lazio	560	694	80.69
Liguria	261	300	87.00
Lombardia	1680	1926	87.23
Marche	304	343	88.63
Molise	131	163	80.37
Piemonte	1330	1453	91.53
Puglia	395	498	79.32
Sardegna	445	513	86.74
Sicilia	506	658	76.90
Toscana	444	515	86.21
TrentinoAlto Adige	298	330	90.30
Umbria	130	170	76.47
Valle d'Aosta	76	85	89.41
Veneto	737	839	87.84
Total	9338	10967	85.15

Area1	Area2	M_1	M_2	W	p value	t	p value
North	South	0.892	0.797	64	0.0001	6.2994	0.0001
North	Centre	0.892	0.830	29	0.0283	2.1984	0.1073
Centre	South	0.830	0.797	22	0.3677	1.0855	0.3314

Source own elaboration. Data retrieval as of 24th January 2016 from <http://www.magellanopa.it/bussola/page.aspx?s=cruscottobussola&q=I/LavitAEpUzr|ZVifNgwQ>

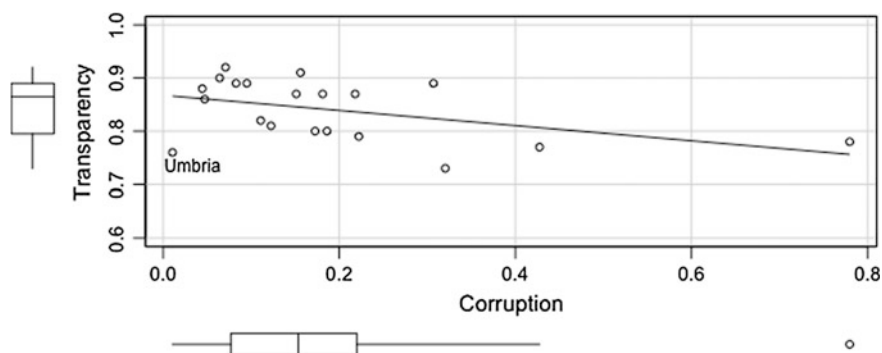
**Fig. 13.1** Corruption and transparency Source own elaboration

Table 13.4 Correlation analysis

	Estimate	SE	<i>t</i> value	<i>Pr</i> (> <i>t</i>)	Multiple <i>R</i> ²	Pearson's <i>r</i>
Correlation analysis 1: CI 2012 vs. TI 2016: 20 regions						
Intercept	0.86744	0.01751	49.53	***	0.192	-0.438
CI	-0.14288	0.06908	2.068	*		
Correlation analysis 2 CI 2012 vs. TI 2016): 19 regions (Umbria excluded)						
Intercept	0.88026	0.01645	53.500	***	0.3214	-0.567
CI	-0.17949	0.06325	2.838	*		
Correlation analysis 3: CI 2016 (estimated) vs TI 2016: 20 regions						
Intercept	0.86806	0.01784	48.65	***	0.189	-0.435
CI	-0.14495	0.07072	-2.05	*		
Correlation analysis 4: CI 2016 (estimated) vs. TI 2016: 19 regions (Umbria excluded)						
Intercept	0.88220	0.01672	52.759	***	0.3307	-0.575
CI	-0.18718	0.06453	-2.909	**		

The significance level is marked with “***”, “****” corresponds to a *p* value between zero and 0.01; “**” corresponds to a *p* value between 0.01 and 0.05; “*” corresponds to a *p* value between 0.05 and 0.10

and quality of information provided, the index reflects the consideration given to the transparency issue and the commitment to its implementation by public administrations. In particular the lack of compliance to a basic legal requirement such as the one considered in this analysis, suggests that a negative attitude towards transparency is present. Moreover, transparency data observed in the study are consistent with those of a recent research (AGE.N.AS 2015) on transparency in the Health Care Sector, conducted by the Italian Agency for Health Care Regional Services (Agenzia Nazionale per i Servizi Sanitari Regionali—AGENAS). This agency released in 2015 the first report on the compliance of health care sector organisations to the transparency legislation and in particular to the Transparency Decree. Data was collected in 2015 among 240 health care organisations and consistently with our analysis, they show that when considering the compliance rate in the regional sub-areas, northern and central regions are those which obtain higher results (AGE.N.AS 2015, p. 55) (Table 13.4).

13.5 Discussion and Conclusions

Corruption is a widespread phenomenon whose causes and consequences have been widely investigated in the social science literature. As of the mid 90s, when Transparency International started publishing its worldwide surveys on the perception of corruption, many studies based on international comparisons have been published on this topic. Together with the CPI index other corruption indicators have been developed by both NGOs, for instance the World Bank, as well as private organisations. These survey based indicators have been subject to criticism with

regard to their reliability and adequacy (see, e.g.:Golden and Picci 2005; Knack 2006; Kurtz and Schrank 2007) for a number of different reasons. In particular Treisman (2007) argues that subjective indexes do not actually measure corruption frequency but rather inferences made on the basis of conventional understandings of corruption's causes (Treisman 2007, p. 213). Moreover Seligson (2006) emphasises that national aggregate measures of corruption assume that the same corruption is uniformly spread within each country, thus neglecting the existence of specific areas where this phenomenon is more pervasive. As corruption occurs in specific contexts and sectors, relying on perceptual aggregate measurements could be misleading, especially for policy making purposes (Heywood and Rose 2014). Thus based on the relevance of considering the different spread of corruption within the different countries, the present research analysed the relation between corruption and transparency, from a "within-country" perspective. The research used the case of the twenty Italian regions to assess whether regions affected by higher levels of corruption are also characterised by a lesser commitment towards transparency and hence by lower levels of compliance to the transparency provisions.

To fulfil the research aim, an objective measure of corruption was used and an index of transparency was developed based on the compliance of public administrations to the transparency provisions issued by the Italian legislator in 2013.

The study found a strong negative correlation between corruption and transparency which seems to suggest a detrimental effect of corruption on public administrations' compliance to the disclosure obligations set in by the transparency legal provisions.

Furthermore the study findings highlight once more the problem of a country divided into two areas, north and south, which is consistent with that depicted in Putnam's study (Putnam et al. 1994). The same results are consistent with those studies on the Italian north-south divide which highlight that in regions with a higher social capital, public administrations are more efficient (Arpaia et al. 2009; Felice and Vasta 2015; Giordano et al. 2009; Nifo and Vecchione 2015b; Pavan et al. 2014; Putnam et al. 1994).

Felice and Vasta (2015) explain this regional divide demonstrating that two different paths of modernisation took place in Italy from its Unification until the first decade of the twenty-first century. According to the authors while in the northern and central regions a process of regional *active modernisation*⁶ occurred—which implies that the actors involved in the modernisation process adopt a rational

⁶Felice and Vasta drawing from the work of Cafagna (1988) developed the concepts of regional active and passive modernisation. According to the authors: "...we have regional active modernization when local elites actively participate to the modernizing process, by sharing common values and coherently implementing the views of the national "historic bloc" (Felice and Vasta 2015, p. 45). Passive modernisation on the contrary occurs when "...there is no "identification" between the elite which advocates modernization and the rest of the community" and on the contrary there are: "..... "extractive" political and economic institutions, where the elites have the interest to pursue some modernization in order to grasp the resulting extra-output, yet preventing the rest of the population from taking any advantage of it" (Felice and Vasta 2015, p. 45).

strategy (Felice and Vasta 2015, p. 45)—in the southern ones *passive modernisation* took place—a process that is not supported by a coherent strategy but, rather, is the result of a passive adaptation pattern—(Felice and Vasta 2015, p. 45). This process—which has its roots in the unwillingness of local political and economical institutions to favour a generalised economic development—led this part of the country into a vicious circle characterised by market failures and unproductive investments mostly resulting from nepotism and illegal activities (Felice and Vasta 2015, pp. 59–60). In such vicious circle corruption finds a fertile ground in the minor economic development which gives rise to clientelism and rents, which in turn foment, as Stiglitz (2002a) argues, the political incentives for information not to be disclosed.

From a policymaking perspective the research results thus highlight, on the one hand, that differing levels of corruption due to different cultural settings necessitate different solutions to the problem (Shim and Eom 2008). On the other, they point out the relevance of carefully considering social and cultural differences when transparency policies are implemented to curb corruption on account of their influence in shaping the related outcomes (Bertot et al. 2010; Brown and Cloke 2004).

From a theoretical perspective the research findings are consistent with Bastida and Benito (2007) who claim that in less corrupt country budget transparency is higher. It emerges, in this respect, the need for additional studies aimed at assessing corruption as a determinant of transparency rather than as its consequence, as the prevailing literature on transparency and corruption does.

Transparency is by all means an important device in the anti-corruption toolbox, but as the study suggests its action might be lessened by the same misbehaviours it tries to contest. As Stiglitz argues there are strong incentives to reduce transparency as secrecy is “*an artificially created scarcity of information*” which generates rents that can be appropriated through illegal activities (Stiglitz 2002a, p. 488).

As the proxy used to measure transparency only provides information about the attitude towards transparency, further studies should be conducted in order to assess whether the same results can be observed when considering transparency actual practices. To this aim a measure of transparency should be developed which allows to satisfy both the publicity and accountability conditions (Lindstedt and Naurin 2010). The former implies that the recipients of the information should be enabled to access and process the same information, while the latter requires that the recipients of the information should also have a power and an incentive to act, through a sanctioning/rewarding system, based on the information provided. Furthermore other variables should be included in the analysis so as to provide more robust results on the relation between corruption and transparency. For instance it could be useful to assess the influence on transparency of variables like citizens' education, the size, type and sector of organisation, as well as measures of institutional quality other than corruption.

References

- AGE.N.AS. (2015). Trasparenza, etica e legalità nel settore sanitario: Primo rapporto sullo stato di attuazione delle azioni adottate dalla sanità pubblica in materia di trasparenza e integrità.
- Andersen, T. B. (2009). E-Government as an anti-corruption strategy. *Information Economics and Policy*, 21(3), 201–210. doi:[10.1016/j.infoecopol.2008.11.003](https://doi.org/10.1016/j.infoecopol.2008.11.003).
- Arndt, C., & Oman, C. (2006). *Uses and abuses of governance indicators*. Paris: OECD Development Centre Studies.
- Arpaia, C. M., Doronzo, R., & Ferro, P. (2009). Computerization, accounting transparency and competitiveness of public administration: A regional analysis. *Bank of Italy Occasional Paper* 48.
- Ball, C. (2009). What Is Transparency? *Public Integrity*, 11(4), 293–308. doi:[10.2753/pin1099-9922110400](https://doi.org/10.2753/pin1099-9922110400).
- Bastida, F., & Benito, B. (2007). Central government budget practices and transparency: An international comparison. *Public Administration*, 85(3), 667–716. doi:[10.1111/j.1467-9299.2007.00664.x](https://doi.org/10.1111/j.1467-9299.2007.00664.x).
- Bauhr, M., & Grimes, M. (2014). Indignation or resignation: The implications of transparency for societal accountability. *Governance*, 27(2), 291–320. doi:[10.1111/gove.12033](https://doi.org/10.1111/gove.12033).
- Bellver, A., & Kaufmann, D. (2005). Transparenting transparency: Initial empirics and policy applications. SSRN 808664.
- Bertot, J. C., Jaeger, P. T., & Grimes, J. M. (2010). Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly*, 27(3), 264–271. doi:[10.1016/j.giq.2010.03.001](https://doi.org/10.1016/j.giq.2010.03.001).
- Bhatnagar, S. (2003). E-government and access to information. *Global Corruption Report, 2003*, 24–32.
- Brown, E., & Cloke, J. (2004). Neoliberal reform, governance and corruption in the south: Assessing the international anti-corruption crusade. *Antipode*, 36(2), 272–294. doi:[10.1111/j.1467-8330.2004.00406.x](https://doi.org/10.1111/j.1467-8330.2004.00406.x).
- Cafagna, L. (1988). Modernizzazione attiva e modernizzazione passiva. *Meridiana*, 2(2), 229–240.
- Campos, J. E., & Pradhan, S. (2007). *The many faces of corruption: Tracking vulnerabilities at the sector level*. Washington, DC: World Bank.
- Commissione Europea. (2014). Relazione della Commissione Europea al Consiglio e al Parlamento Europeo: Relazione dell'Unione sulla lotta alla corruzione. Allegato sull'Italia.
- Curtin, D., & Meijer, A. J. (2006). Does transparency strengthen legitimacy? *Information Polity*, 11(2), 109–122.
- European Commission Directorate-General for Home Affairs. (2013). Special Eurobarometer 397 —Corruption Report.
- Felice, E., & Vasta, M. (2015). Passive modernization? The new human development index and its components in Italy's regions (1871–2007). *European Review of Economic History*, 19(1), 44–66. doi:[10.1093/ereh/heu018](https://doi.org/10.1093/ereh/heu018).
- Gerring, J., & Thacker, S. C. (2004). Political institutions and corruption: the Role of unitarism and parliamentarism. *British Journal of Political Science*, 34(2), 295–330.
- Giordano, R., Tommasino, P., Casiraghi, M., & Cannari, L. (2009). Le determinanti dell'efficienza del settore pubblico: il ruolo della cultura e delle istituzioni. In *Mezzogiorno e politiche regionali* (pp. 253–275). Torino: Edizioni della Sera.
- Golden, M. A., & Picci, L. (2005). Proposal for a new measure of corruption, illustrated with Italian data. *Economics and Politics*, 17(1), 37–75. doi:[10.1111/j.1468-0343.2005.00146.x](https://doi.org/10.1111/j.1468-0343.2005.00146.x).
- Grimmelikhuisen, S., Porumbescu, G., Hong, B., & Im, T. (2013). The effect of transparency on trust in government: A cross-national comparative experiment. *Public Administration Review*, 73(4), 575–586.
- Gupta, S., Davoodi, H., & Alonso-Terme, R. (2002). Does corruption affect income inequality and poverty? *Economics of Governance*, 3(1), 23–45.

- Gupta, S., Davoodi, H. R., & Tiongson, E. (2001a). Corruption and the provision of health care and education services. In A. K. Jain (Ed.), *The political economy of corruption*. London: Routledge.
- Gupta, S., De Mello, L., & Sharan, R. (2001b). Corruption and military spending. *European Journal of Political Economy*, 17(4), 749–777. doi:10.1016/s0176-2680(01)00054-4.
- Gyimah-Brempong, K. (2002). Corruption, economic growth, and income inequality in Africa. *Economics of Governance*, 3(3), 183–209.
- Harrison, T. M., & Sayogo, D. S. (2014). Transparency, participation, and accountability practices in open government: A comparative study. *Government Information Quarterly*, 31(4), 513–525. doi:10.1016/j.giq.2014.08.002.
- Heald, D. (2006). Varieties of transparency. In C. Hood & D. Heald (Eds.), *Transparency: The key to better governance?* (pp. 25–41). Oxford: Oxford University Press/British Academy.
- Heywood, P. M., & Rose, J. (2014). “Close but no Cigar”: The measurement of corruption. *Journal of Public Policy*, 34(03), 507–529.
- IMF. (2015). The IMF and good governance. A factsheet. <http://www.imf.org/external/np/ext/facts/gov.htm>.
- Islam, R. (2002). Into the looking glass: What the media tell and why—an overview. In W. B. Institute (Ed.), *The right to tell: The role of mass media in economic development*. Washington, DC: World Bank, WBI Development Studies.
- Kaufmann, D. (2002). Transparency, incentives and prevention (TIP) for corruption control and good governance. In *Empirical findings, practical lessons, and strategies for action based on international experience*. Washington, DC: The World Bank. http://www.worldbank.org/wbi/governance/pdf/quinghua_presentation.pdf.
- Knack, S. (2006). Measuring corruption in Eastern Europe and Central Asia: A critique of the cross-country indicators. *World Bank Policy Research Working Paper 3968*.
- Kolstad, I., & Wiig, A. (2009). Is transparency the key to reducing corruption in resource-rich countries? *World Development*, 37(3), 521–532. doi:10.1016/j.worlddev.2008.07.002.
- Kurtz, M. J., & Schrank, A. (2007). Growth and governance: Models, measures, and mechanisms. *Journal of Politics*, 69(2), 538–554. doi:10.1111/j.1468-2508.2007.00549.x.
- Lambsdorff, J. G. (2005). *Consequences and causes of corruption: What do we know from a cross-section of countries?* Passauer Diskussionspapiere: Volkswirtschaftliche Reihe.
- Lindstedt, C., & Naurin, D. (2010). Transparency is not enough: Making transparency effective in reducing corruption. *International Political Science Review*, 31(3), 301–322.
- Lio, M. C., Liu, M. C., & Ou, Y. P. (2011). Can the internet reduce corruption? A cross-country study based on dynamic panel data models. *Government Information Quarterly*, 28(1), 47–53. doi:10.1016/j.giq.2010.01.005.
- Mauro, P. (1995). Corruption and growth. *The Quarterly Journal of Economics*, 110(3), 681–712.
- Mauro, P. (1998). Corruption and the composition of government expenditure. *Journal of Public Economics*, 69(2), 263–279.
- Meijer, A. J. (2013). Understanding the complex dynamics of transparency. *Public Administration Review*, 73(3), 429–439. doi:10.1111/puar.12032.
- Meijer, A. J., Curtin, D., & Hillebrandt, M. (2012). Open government: Connecting vision and voice. *International Review of Administrative Sciences*, 78(1), 10–29. doi:10.1177/0020852311429533.
- Michener, G., & Bersch, K. (2013). Identifying transparency. *Information Polity: The International Journal of Government & Democracy in the Information Age*, 18(3), 233–242. doi:10.3233/ip-130299.
- Montinola, G. R., & Jackman, R. W. (2002). Sources of corruption: A cross-country study. *British Journal of Political Science*, 32(1), 147–170.
- Nifo, A., & Vecchione, G. (2014). Do institutions play a role in skilled migration? The case of Italy. *Regional Studies*, 48(10), 1628–1649. doi:10.1080/00343404.2013.835799.
- Nifo, A., & Vecchione, G. (2015a). Institutional Quality Index—Version III.
- Nifo, A., & Vecchione, G. (2015b). Measuring institutional quality in Italy. *Rivista Economica del Mezzogiorno*, 1–2(2015), 157–182.

- OECD. (2014). Boosting integrity and fighting corruption at the OECD.
- Otenyo, E. E., & Lind, N. S. (2004). Faces and phases of transparency reform in local government. *International Journal of Public Administration*, 27(5), 287–307. doi:10.1081/pad-120028811.
- Pallot, J. (2001). Transparency in local government: Antipodean initiatives. *European Accounting Review*, 10(3), 645–660. doi:10.1080/09638180126794.
- Pavan, A., Reginato, E., & Fadda, I. (2014). *The implementation gap of NPM reforms in Italian local governments. An empirical analysis*. Milano: Franco Angeli.
- Plummer, J., & Cross, P. (2006). Tackling corruption in the water and sanitation sector in Africa. In J. E. Campos & S. Pradhan (Eds.), *The many faces of corruption* (pp. 221–263). Washington, DC: World Bank.
- Putnam, R. D., Leonardi, R., & Nanetti, R. Y. (1994). *Making democracy work: Civic traditions in modern Italy*. Princeton, NJ: Princeton University Press.
- Relly, J. E., & Sabharwal, M. (2009). Perceptions of transparency of government policymaking: A cross-national study. *Government Information Quarterly*, 26(1), 148–157.
- Reynaers, A.-M., & Grimmelhuisen, S. (2015). Transparency in public-private partnerships: Not so bad after all? *Public Administration*, 93(3), 609–626. doi:10.1111/padm.12142.
- Seligson, M. A. (2006). The measurement and impact of corruption victimization: Survey evidence from Latin America. *World Development*, 34(2), 381–404.
- Shim, D. C., & Eom, T. H. (2008). E-government and anti-corruption: Empirical analysis of international data. *International Journal of Public Administration*, 31(3), 298–316. doi:10.1080/01900690701590553.
- Stiglitz, J. E. (2002a). Information and the change in the paradigm in economics. *The American Economic Review*, 92(3), 460–501.
- Stiglitz, J. E. (2002b). Transparency in government. In World Bank Institute (Ed.), *The right to tell: The role of mass media in economic development*. Washington DC: World Bank: WBI Development Studies.
- Tanzi, V. (1998). Corruption around the world: Causes, consequences, scope, and cures. *Staff Papers-International Monetary Fund*, 45(4), 559–594.
- Transparency International. (2015). Corruption Perceptions Index 2015. https://www.transparency.it/wp-content/uploads/2016/01/Corruption-Perceptions-Index-2015-report_EMBARGO.pdf.
- Treisman, D. (2007). What have we learned about the causes of corruption from ten years of cross-national empirical research? *Annual Review of Political Science*, 10(1), 211–244. doi:10.1146/annurev.polisci.10.081205.095418.
- United Nations. (2003). United Nations Convention Against Corruption.
- Weisberg, S. (2014). *Applied linear regression*. Hoboken, NJ: Wiley.
- Welch, E. W., Hinnant, C. C., & Moon, M. J. (2005). Linking citizen satisfaction with e-government and trust in government. *Journal of Public Administration Research and Theory*, 15(3), 371.
- Williams, A. (2015). A global index of information transparency and accountability. *Journal of Comparative Economics*, 43(3), 804–824. doi:10.1016/j.jce.2014.10.004.

Chapter 14

Performance Measurement Systems in Universities: A Critical Review of the Italian System

Natalia Aversano, Francesca Manes-Rossi and Paolo Tartaglia-Polcini

Abstract This research focuses on the development of performance measurement systems (PMS) in universities and discusses a possible move toward international harmonization. To this end, the case of Italy is examined. In particular, the new regulation issued by the national Agency (ANVUR) is examined and compared with the guidelines provided by the IPSASB, in the aim of shedding light on the gap between the Italian guidelines and the approach proposed at international level by the IPSASB. IPSASB guidelines represent a good reference for a harmonized PMS across EU member-States. Results evidence that the Italian guidelines fall short of the approach followed on the international scene, reducing international comparability and transparency.

Keywords Performance measurement · Universities · Italy · ANVUR guidelines · IPSAS RPG 3

14.1 Introduction

There has long been an interest in performance measurement in the public sector domain (Beyle and Parratt 1938). However, with the advent of New Public Management reforms (Hood 1995; Lapsley 2008) the need for implementing Performance Measurement Systems (PMS) has become imperative as a pivotal tool for achieving efficiency and effectiveness, for supporting evaluation processes, as

N. Aversano (✉)

Department of Mathematics, Computer Science and Economics, University of Basilicata,
Potenza, Italy
e-mail: natalia.aversano@unibas.it

F. Manes-Rossi · P. Tartaglia-Polcini

Department of Management & Innovation Systems, University of Salerno, Salerno, Italy
e-mail: fmanesrossi@unisa.it

P. Tartaglia-Polcini

e-mail: ptpolcini@unisa.it

© Springer International Publishing AG 2018

E. Borgonovi et al. (eds.), *Outcome-Based Performance Management
in the Public Sector*, System Dynamics for Performance Management 2,
DOI 10.1007/978-3-319-57018-1_14

well as for enhancing accountability in public entities (de Lancer Julnes and Holzer 2001; Poister 2003; van Dooren and van de Walle 2008; van Helden et al. 2012).

Following Broadbent and Laughlin (2009), within the public domain, public universities represent a kind of organization in which a PMS is particularly necessary in order to improve public efficiency and effectiveness and support the decision-making of public administrators. In fact, universities have traditionally included the management and transmission of knowledge as their primary objectives, objectives that are difficult to measure. Moreover, increased financial pressures and related budget cuts in recent years have heightened the need for effectiveness and efficiency, and international competition has played a significant role in managing higher education institutions, resulting in a more important role for PMS (Higgins 1989).

PMS are generally designed taking into account the peculiarity of the specific organization; however, a tendency to identify some fundamental elements in order to favour comparison and benchmarking among public sector entities has emerged between standard setters and international organizations (OECD 2007; GASB 2010; etc.). To promote a harmonized approach to performance measurement in public entities, in 2015, after a long consultation process that started in 2011, the International Public Sector Accounting Standards Board (IPSASB) issued the Recommended Practice Guideline n. 3 (RPG 3) entitled “Reporting Service Performance Information”. This RPG 3 provides guidance on reporting service performance information in General Purpose Financial Reports (GPFs). Following the IPSASB approach, Service Performance Information is information about the services that the entity provides, its service performance objectives and the extent of its achievement of those objectives, and it is aimed at assisting users of GPFs to assess the entity’s service efficiency and effectiveness (Manes Rossi and Aversano 2015).

Notwithstanding, differences in the adoption of performance measurement and reporting systems worldwide may result as a consequence of the differences in the institutional, cultural and historical contexts of each country (Benito et al. 2007). Scholars have underlined the need for greater attention to PMS at a macro-level (Modell 2003).

The present research aims to analyze the PMS designed by the Italian National Agency for the Evaluation of the University and Research Systems (ANVUR) in order to shed light on the gap between the Italian guidelines and the approach proposed at international level by the IPSASB.

The analysis of the degree of the adoption of the RPG 3 in a European country—such as Italy—shows to what extent IPSASB guidelines may constitute a good point of reference for a harmonized PMS, improving transparency and comparability throughout EU member states.

The Italian case has been selected as an example of a university system in which numerous reforms have been undertaken that have profoundly changed organization, management, budgeting and accounting systems of universities, including the mandate adoption of a PMS. However, a bureaucratic approach seems to persist in the use of performance data.

To meet the research objective, a literature review noting the evolution of studies on PMS, with special regard to the public sector and more specifically to universities, is provided in the second section. The third section introduces the Italian context and the content of the RPG 3. In the fourth section, the two sets of guidelines are examined with the aim of comparing the structure and the content as well as exploring the reasons for the differences. The fifth section concludes the research, showing that the Italian guideline presents a different structure with respect to those provided by the IPSASB and follows an institutional approach with a focus on indicators, objectives and actors. Furthermore, the IPSASB's RPG 3 provides basic definitions and principles for the "Presentation and Organization of Service Performance Information". Limitations and further developments of the research are also addressed.

14.2 The Evolution of Performance Measurement Systems in Universities: A Literature Review

Scholars have largely investigated PMS in the public sector domain, identifying different concepts and attributing different characteristics to these systems; all the scholars agree that PMS include metrics attuned to measuring the efficiency and effectiveness of actions undertaken in an organization (Broadbent and Laughlin 2009; Cuganesan et al. 2014; Ferreira and Otley 2009; Modell 2003; Poister 2003; van Dooren et al. 2015; Yang and Holzer 2006). Given the focus of the research on the elements of the PMS designed by the ANVUR for Italian universities in comparison with the IPSASB's requirements, it is worth clarifying that, in accordance with Ferreira and Otley, "we see this term as including all aspects of organizational control, including those included under the heading of management control systems" (Ferreira and Otley 2009, p. 264). From this perspective, PMS are a kind of integrated technical system able to collect and provide information for decision-making in order to elaborate on achievable goals, encourage interaction among members of the organization and motivate the employees (Esposito et al. 2013). However, PMS are also aimed at satisfying the demand for a wide accountability (Kloot and Martin 2000; Sanderson 2001; Broadbent and Laughlin 2009; Barrados and Blain 2013).

Scholars have discussed implementation difficulties in adopting PMS in public sector organizations, highlighting the need for a multidimensional perspective able to tackle the complexity of this kind of entity (Broadbent and Laughlin 2009; Ferreira and Otley 2009). Moreover, the unintended consequences of performance assessments and the riskiness of public sector performance measurement have been also discussed at length (a synthesis in Van Thiel and Leeuw 2002; Cuganesan et al. 2014) and can offer a lesson to learn while discussing the features of new systems: measurement errors as well as problems concerning the content, position and amount of measures (Bouckaert and Balk 1991); the possible "ossification" effect

related to the adoption of performance indicators that can impede or limit innovation (Smith 1995); the risk of symbolic behaviour as a consequence of monitoring (i.e. monitoring seems to be applied but it is not) (Van Thiel and Leeuw 2002) or even a possible distortion of performance information (van Dooren et al. 2015). In this vein, Pollitt (2013) suggests paying due attention to the logic used by different actors approaching PMS in public sector entities.

A further point addressed by previous studies refers to the wide range of users of performance measures: politicians, managers and citizens may all be interested in performance measures even if from different points of view (Greiling 2005; Johnsen 2005; van Dooren et al. 2015). In this regard, Behn (2003) has emphasized the need to adopt different measures for different purposes: citizens are interested in evaluating program effectiveness; managers may use performance measures in the budget formulation as well as to detect areas for improvement (thus both for evaluation and control) and for motivating employees; politicians may consider these measures both for public accountability and decision-making and even for internal comparison.

In order to play a significant role in improving citizens' trust in public entities, "performance measurement should become a tool that can facilitate double-loop learning in the never-ending pursuit of excellence" (Yang and Holzer 2006, p. 123).

While numerous studies report on the experiences of different public organizations implementing PMS (Cavalluzzo and Ittner 2004; Ammons and Rivenbark 2008; Torres et al. 2011; Bianchi and Rivenbark 2012; Brusca et al. 2015) and on how PMS modernize entities (Sotirakou and Zeppou 2006), publications concerning the adoption of these systems, their features and their limits in the context of higher education are quite few (Modell 2003; Arnaboldi and Azzone 2010; Kuah and Wong 2011; ter Bogt and Scapens 2012; Esposito et al. 2013; Christopher and Leung 2015; Kallio et al. 2016). Some previous works discuss the research assessments and the adoption of journal ranking as a tool to evaluate research or concentrate the attention on PMS as a tool for governance and resource allocations (as reported in ter Bogt and Scapens 2012). Lapsley and Miller (2004) have highlighted that all reforms implemented in universities need to be examined in terms of the political context.

It is widely recognized that a PMS assumes a fundamental role in organizations, such as universities, where inputs and outputs are mainly intangible and outcomes are notoriously difficult to measure (Cosenz 2011). Since this research focuses on this realm, it is worth examining both the evolution of PMS in the international context and results reported by previous research with the aim of also understanding the widespread adoption of these systems at an international level.

Under the NPM wave, the number of public universities adopting PMS increased, basically as a consequence of the changes in the way in which universities can fund their activities and because of the reduction in public funds.

King Alexander reports that by 1998 several universities in the United States had already adopted performance funding and performance budget (King Alexander 2000). In Australia, the Council of Australian Governments developed a framework for all public organizations involved in human service provision (education, health,

housing, etc.) and, while recognizing the risks and limitations connected with performance indicators, suggested programme and operational indicators attuned to control efficiency and effectiveness in relation to a number of dimensions in the aim of improving accountability (Van Thiel and Leeuw 2002). In Europe, the so-called Bologna Process introduced “the qualification frameworks” (Bologna Working Group 2005), and a stronger emphasis has been placed on students’ learning outcomes (OECD 2009). According to the European Consortium for accreditation (ECA), “Data collection and development of performance indicators should strictly adhere to the principles of transparency, readability and accountability of European education, thus allowing for measuring and comparing the strengths of institutions” (ECA 2009, p. 3). However, more recent literature is also critical of the unintended consequences of a performance management approach in universities, especially in terms of the assessment of the individual performance of researchers (ter Bogt and Scapens 2012; Kallio et al. 2016).

In the 1980s in the UK, the Jarratt Committee focused attention on the need to adopt a PMS and evaluated the performance of individual academics (Jones 1991). To date, the UK universities have shifted from historical budgets to block grants assigned in accordance with performance, where the elements for research are based on research ratings (ter Bogt and Scapens 2012). Similarly, in the Netherlands, reforms implemented in recent decades have focused on relating funds to performance and are basically output-oriented (Sousa et al. 2010; ter Bogt and Scapens 2012). In the Nordic countries, quality assessment programmes have been implemented, bringing to light the assessment of learning outcomes. In Norway, the model is basically process-oriented and considers indirect measures, such as the quality assurance system. In Sweden, it focuses on the students’ learning outcomes and on courses (Pettersen 2014). In the Finnish universities, management by results was implemented in the late 1980s when universities’ budgets started to include performance-based funds, and in 2006 a highly structured performance-related pay system was introduced; it has been highly criticized by scholars (Kallio et al. 2016).

It is worth noting that a peculiar system of performance evaluation was introduced in Austria in 2002: together with the development of the performance contract and the performance report publication, an Intellectual Capital report has to be published by each university (Leitner 2004).

While scholars have discussed experiences in different countries, to the best of our knowledge, a lack of attention has been paid to a possible harmonization of performance measures among universities. The IPSASB has made some efforts to create some guidelines to clarify what a service performance report should include and how it should be organized by public reporting entities. This research attempts to analyze the performance measurement systems designed by the ANVUR in order to shed light on the gap between the Italian guidelines and the approach proposed at an international level by the IPSASB.

14.3 The National and International Context

14.3.1 *The Italian Context*

The Italian university system has been based on a strongly bureaucratic model since the 1980s. Universities were previously almost completely state-owned and managed in a centralized way. The government decided how many resources each university might receive and how these resources had to be expended (the budget was distributed between subject areas and expenditure lines by the central government). No evaluation system of the effectiveness and efficiency of university activities was implemented.

In the mid-1990s, the NPM forced a rethink of the traditional national university system and urged a shift from a bureaucratic model to a managerial model, implementing measures for effectiveness and efficiency.

The first step was to provide universities with a certain degree of autonomy (Law n. 168/1989, “Legge Ruberti”). The central government decided the number of resources to transfer to each university and the general rules to be followed in expending them. Each university had, for the first time, the freedom to choose how to manage its resources, aware that its results would be measured and evaluated by the central government (Ministry).

Law n. 537/1993 created a central national body (the *Osservatorio Nazionale per la Valutazione del Sistema Universitario*, OVSU, National Observatory for the University System) with the aim to monitor and register the activity of each university. In addition, an internal committee (*Nucleo di Valutazione*, Internal Evaluation Committee) was created with the task of verifying the efficient use of public resources by each university. In 1999, Law 370/1999 transformed the OVSU into CNVSU (*Comitato Nazionale per la Valutazione del Sistema Universitario*, National Committee for the University System Evaluation), a national committee responsible for monitoring the performance of universities. The same law emphasized the role of the internal evaluation committees of each university in providing the CNVSU with information and data collected during the performance of their duties.

For the first time, significant attention was paid to the evaluation of academic activities. In order to implement a common model for performance evaluation, the CRUI (*Conferenza dei Rettori delle Università Italiane*, Conference of Rectors of Italian Universities, the main federation of universities in Italy) launched a project called “Campus One”, aimed at developing and implementing concepts and methods of performance evaluation in the academic world. In 1998, the Decree n. 204 implemented the Research Evaluation Committee (CIVR, *Comitato di Indirizzo per la Valutazione della Ricerca*, Committee for Research Evaluation), whose goal was to evaluate the research activity of universities and to promote improvement. Therefore, the academic performance central evaluation system was composed of the CNSVU for the teaching activity and the CIVR for the research activity; the local evaluation of universities was performed by the internal evaluation

committees (so-called *Nuclei di Valutazione*), whose importance would become increasingly great over time. The internal evaluation committees, on the other hand, supported the governance of their university, providing it with systematic informational flows and analysis about relevant aspects of the performance.

In the first half of 2000s, the culture of evaluation became more and more widespread, the evaluation activity became systematic and databases were created to provide the Ministry and each university with information on decisions. The basic ideas for this development started from the Bologna Process, formalized in 1999 with the Declaration of Bologna. The main concept was the absolutely strategic centrality of a quality assurance system: each country should have its own national system that would have to monitor the internal quality system of each university. On the other hand, at an international level, national quality assurance systems should harmonize with each other on the same basic concepts. These concepts were formalized by the European Association for Quality Assurance in Higher Education (ENQA) in the Standards and Guidelines for Quality Assurance in the European Higher Education Area (ESG). The process of harmonization of quality assurance systems has continued with the creation of the European Register of Quality Assurance Agencies (EQAR) in which the national assurance quality agencies that respect the ESG might be registered.

In 2003, the first systematic evaluation of research activity and output was conducted by the CIVR: the VTR (*Valutazione Triennale della Ricerca*, Evaluation of Research Activity of Three Years: 2001–2003) by which universities were classified according to the research results gained.

In the same period, a culture of self-evaluation (especially of teaching) began to spread hand-in-hand with the increased awareness of the importance for each institution to preserve its efficiency and effectiveness in order to guarantee its own survival in the long run. As a consequence, local internal quality assurance systems became more and more central in the national university system.

Law n. 286/2006, following a model already implemented in other developed countries, established an independent agency, the ANVUR (*Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca*, University and Research Evaluation National Agency) that took the place of CNSV and CIVR. Until 2010, the new agency could not operate effectively for many complex reasons, including a reluctance to evaluate the academic system. In this period, CNSVU and CIVR continued to exist and operate but their real power was greatly reduced. These conditions in Italy were in strong contrast with the wide dissemination of evaluation culture and tools in other European countries.

At the end of 2010, a very important law (Law 240/2010, “Legge Gelmini”) radically reorganized the university system; it was based on a strict relation between resources granted to individuals and institutions and their own performance indicators. The ANVUR became central to the whole system, and it became very active in performing its institutional tasks. The first task performed was the VQR 2004–2010 (*Valutazione della Qualità della Ricerca 2004–2010*, Evaluation of Research Activity of the Years 2004–2010). Another very important and strategic task performed by the ANVUR in 2012/2013 was the procedure of the National Academic

Qualification in which researchers' careers were evaluated in order to grant the qualification of full and associate professorships. The most recent task performed by the ANVUR is the VQR 2011–2014, which is still in progress.

The basic concept on which the Italian evaluation model is based has increasingly shifted from collecting information and data only for external analysis and the evaluation of the performance to a mix of external and internal use. Self-evaluation has become more and more important, based on the idea that a university will implement good practices if they can understand the weaknesses of their own mechanisms and procedures.

Another basic concept is the relevance of the (internal and external) transparency of the performance measures: external and internal recipients of the informational flows will steer the university in virtuous directions if they have the relevant information about performance measurement.

At the moment, the following aspects of the Italian university system may be highlighted:

1. the central agency (ANVUR) defines, especially for teaching, the minimum requirements but does not (despite the principles affirmed) investigate the specific quality of the internal evaluation system of each university;
2. as a consequence of the previous point, universities have mostly quantitative parameters and very little attention is paid to qualitative aspects, especially in teaching (e.g. the teaching methods);
3. students play an insignificant role in the evaluation process; it mainly comprises the completion of questionnaires about their satisfaction with the teaching activity;
4. the external stakeholders (including the private and public organizations as potential employers of graduates) play a minimal role in defining objectives and strategies of universities.

The ANVUR has become increasingly active in improving the national evaluation system, introducing new criteria in designing and performing a university evaluation system more consistent with the principles of the Bologna Process and of ESG, which emphasize qualitative aspects.

14.3.2 The International Context

In March 2008, the International Federations of Accountants (IFAC) through the IPSASB launched a new project on reporting service performance information.

Public sector entities worldwide provide a huge variety of services and operate under different legislative requirements and reporting frameworks; moreover, they have varied levels of experience with such reporting. Therefore, in the international scenario there is little coordinated information available about the ways in which public sector entities set financial and non-financial objectives, measure performance, and report on results.

However, even if there are no identical service performance reporting frameworks at a global level, similarities in the service performance information that are reported have been identified. Starting from these similarities and after the revision and the comparison of existing national standards, guidance, and regulatory requirements for service performance reporting from selected national jurisdictions and some international organizations (UN and OCSE), IPSASB developed and issued a Consultation Paper (CP) titled *Reporting Service Performance Information* in 2011.

Following a traditional pattern for the implementation of new principles and standards, the IPSASB has solicited a discussion on four Preliminary Views (PVs) and five Specific Matters for Comments (SMCs) with the aim to propose: (a) a principles-based framework for reporting service performance information, and (b) its terminology and associated definitions.

After analysis of the responses received from several countries and several categories of respondents, IPSASB decided that information on service performance should be addressed through the development of a Recommended Practice Guideline (RPG) and, on that basis, in December 2013 a draft of RPG was developed with the ED 54, *Reporting Service Performance Information*.

Consequently, on 31 March 2015, the IPSASB published RPG 3 *Reporting Service Performance Information* (IPSASB 2015) that provided guidance for entities planning to start their report of service performance information.

14.4 The National and International Guidelines

In order to shed light on the gap between the Italian guidelines and the approach proposed at international level by the IPSASB, the ANVUR guideline and the RPG 3 of IPSASB are analyzed in the following sections.

14.4.1 *The Italian National Agency for the Evaluation of the University and Research Systems Guideline*

As already outlined, the Italian Ministry of Education, Universities and Research delegated the ANVUR to manage the evaluation system of administrative activities of universities and research institutions. In May 2015, the ANVUR issued guidelines for the integrated management of the performance cycle of Italian state universities. This document is composed of three sections: the first illustrates the framework within which this document has been issued, the second evidences the management tools of the performance cycle and the third indicates the actors of the evaluation system. The guideline often refers to previous legislation and rules issued about performance evaluation for the public sector.

In the first section, ANVUR indicates an important premise, often recalled throughout the document: in order to obtain an effective implementation of integrated performance measurement system, the activities of the administrative and teaching staff should not be considered in a separate way but in an integrated manner in all the phases of the performance management cycle. To this end, ANVUR evokes the main university missions of teaching and conducting scientific research, in addition to the so-called “third mission”. The third mission refers to the social and economic benefits arising from the impact that research and teaching activities have on populations involved (for example on other researchers, students and other members of society).

The ANVUR guideline refers to the framework defined by the Legislative Decree no. 150/2009 according to which, the performance cycle is divided into three phases: (1) programming (the plan of performances); (2) monitoring activity and possible programming correction and (3) an evaluation of the results reported in the “report on performance”.

Within this framework, the guideline tries to change the approach by which governments prepare such documents, looking for integration of all planning required documents as regards to the administrative performance.

To this end, a first important management tool is the *Integrated Plan*, which is analyzed in the second section of the guideline.

The Integrated Plan is a unique document that systematically develops the planning of the administrative activities related to performance, transparency and anti-corruption, taking into account both the strategy for the institutional activities and the economic and financial planning.

The logic behind the Integrated Plan is characterized by an “extended” notion of performance, understood as the ability of universities to interact dynamically with the other entities (both private and public) in a mutually beneficial durable and sustainable relation. The guideline outlines the five main sections of the Integrated Plan, which must be understood as a list of minimum content (see Table 14.1).

The objectives evidenced in the second section must be relevant, pertinent, specific, measurable, oriented to improvement and connected with the resources required.

The guideline shows that the construction of quantitative indicators supports the assessment of the cycle of the performance as they provide the essential empirical basis for performance measurement, within a more complex evaluation process. Accordingly, a list of indicators related to research, teaching and the third mission is provided, basically replicating the set of indicators developed by ANVUR within the VQR and AVA procedures.

The VQR indicators evaluate the quality of research produced by the professors of the departments; they are the final object of the evaluation process. Examples of indicators are the proportion of inactive researchers or the percentage of excellent scores.

These indicators take into consideration both quantitative and qualitative aspects of the research; in fact, both the number of publications and the position of the journal in the scientific journal rankings are considered. The two types of

Table 14.1 The sections of the integrated plan

Sections	Content
I. “ <i>Strategic framing of the university</i> ”	Summarizes the main lines of development, making explicit reference to existing strategic planning documents, financial planning (three-year and annual), and to policies for quality assurance undertaken following the initiation of the AVA system (<i>Autovalutazione, Valutazione periodica, Accreditemento, Self-assessment, Periodic assessment, Accreditation</i>)
II. “ <i>Organizational performance: the framework of action of the General Director</i> ”	Explains <ul style="list-style-type: none"> – the list of <i>objectives</i> that underpin the operations planned; – the <i>indicators</i> chosen for monitoring the actions and measurement of targets; – <i>subjects</i> involved in the actions and that are responsible for the achievement of the objectives
III. “ <i>Analysis of the risk areas</i> ”	Indicates the risk areas of corruption and the consequences of failure or insufficient administrative transparency
IV. “ <i>Communication and transparency</i> ”	Presents the university’s communication strategy and focuses on changes and innovations set in the previous year and/or planned for the upcoming one
V. “ <i>The individual performance: the performance measurement systems and incentives</i> ”	Describes the logic and the method used for the definition and assignment of individual goals, especially as regards monetary incentives

assessment should be used in a balanced way. The quality aspects are declared as relevant, even if they often remain secondary to the importance of quantitative objectives. As Kallio and Kallio (2014) noted, the use of only quantitative measures is not sufficient since they are not reflective of the quality of research.

The AVA indicators analyze the educational processes. Examples of indicators are the number of students who leave the course, the duration of the degree course, student satisfaction and employment opportunities. The ANVUR monitors the parameters and indicators for accreditation and periodic evaluation for the allocation of rewards to universities. However, this set of indicators is mainly quantitative.

The indicators relating to the third mission measure the results at university level and refer to the valorisation of the research (examples of indicators are spin-off number and patent number) and of the social and cultural enhancement (such as public engagement initiatives or visitors of museum).

Concerning the indicators of administrative performance, the performance cycle of administrative staff should include a set of detailed indicators. They are individually assessed on the basis of their activities to support the education and research activities. Possibly, the objectives could be defined and measured by a *Management*

by *Results* (MBR) model in order to motivate and empower administrative employees to work according to the university's strategy (Kallio and Kallio 2014).

The link between the objectives of the performance cycle and institutional objectives of teaching, research and their repercussions in the form of the third mission should be made explicit. The indicators maximize their utility when they are comparable with organizations with similar characteristics (institutional mission, size, geographic location, etc.).

The organizational units (departments) are responsible for achieving the objectives indicated by the normative framework for the institutional activities of the universities, while the General Director is responsible for the administrative activities.

In addition to the integrated plan, the other management tool analyzed by the guideline is the "*Report of the Performance*". The university draws up this document to report on organizational and individual results achieved in the previous year. The report thus provides a balanced reflection on the results achieved by the administrative staff.

In terms of the *actors* of the evaluation system, the guideline indicates that it is driven by the ANVUR and fuelled by the activity of the evaluation groups aimed at ensuring high-quality university activities.

The ANVUR guideline evidences a PMS based on a managerial approach and focused on results. The Italian guideline aims to develop a performance measurement system that considers, in an integrated way, the activities of the administrative and teaching staff. However, even if the integrated approach can favour the motivation and the achievement of the objectives as well as the interaction of the employees (Esposito et al. 2013), the two different approaches should be followed for the measurement of administrative and teaching staff performance.

The academic staff members that conduct research and teaching activities should be evaluated using both quantitative and qualitative indicators of equal importance. Moreover, it should be noted that monetary compensation is an important aspect for the motivation and rewarding of academic staff, even though job satisfaction is often more important (Kallio and Kallio 2014).

The evaluation of the administrative staff should be based mainly on the quantitative performance indicators based on efficiency and effectiveness measures. In this case, the monetary compensation represents the most important factor for the motivation and the responsibility.

14.4.2 Recommended Practice Guideline 3 Reporting Service Performance

The RPG 3 is composed of several sections that reflect the objective of the document, the meaning of some terms, the reporting boundary, the annual reporting and reporting period, the principles for presentation, the selection, the location and the

organization of service performance information and the basis for conclusions and some illustrative examples.

This RPG 3 provides guidance on reporting service performance information in General Purpose Financial Reports (GPFRs). Reporting Service performance information is important to allow users to evaluate the services provided by public sector entities and to assess the entities' efficient and effective use of resources. Considering the differences of services, frameworks and reports in the international context, the RPG 3 should not standardize service performance reporting, but should aim at providing guidelines on good practice and requirements. This RPG 3 is applicable to all public sector entities and it is intentionally non-prescriptive; IPSASB only encourages its application by suggesting that reporting of information in accordance with this RPG 3 represents a good practice.

The RPG 3 defines seven of the most important terms: *Effectiveness*, *Efficiency*, *Inputs*, *Output*, *Outcome*, *Performance indicators* and *Service performance objective*. Moreover, the implementation examples that accompany RPG 3 illustrate the terms defined above. In particular, it is evidenced that service performance objectives may be expressed using a narrative description of a desired future state resulting from provision of services or using performance indicators of inputs, outputs, outcomes or efficiency, or through a combination of one or more of these four performance indicators. Moreover, service performance objectives will generally be specific, measurable, achievable, realistic and time-bound.

Concerning the *reporting boundary*, service performance information should be reported at least annually and cover the same reporting period as that covered by the financial statements; however, the reporting period may be different in consideration of users' needs.

As regards the *Principles for the Presentation of Service Performance Information*, the RPG 3 evidences that an entity should present service performance information that is useful to users for accountability and decision-making purposes. Moreover, if the service performance information is presented in combination with the information in an entity's financial statements, users should be able to assess the entity's finances in the context of its achievement of service performance objectives and vice versa.

The presentation of service performance information should achieve the qualitative characteristics described by the Conceptual Framework. The aggregation or disaggregation of service performance information should be at a level that increases the transparency and the understandability. The comparability to other entities can be difficult to achieve in the context of service performance information since diverse services are provided and, in the case of the same services provided, the objectives can be different.

The overriding principle regarding *indicators* is that they should be selected based on their importance to users and their usefulness in assessing the entity's achievements in terms of its service performance objectives. Moreover, to avoid overwhelming users, entities generally will need to identify only those few key performance indicators that will best meet the needs of users for information that, in turn, meets the objectives of financial reporting.

IPSASB indicates that the *location of Service Performance Information* may be presented either as part of a GPFR that includes the financial statements or in a separately issued GPFR. To make this decision several factors should be considered:

- whether the needs of users and the qualitative characteristics would be enhanced if the service performance information is included in the same GPFR as the financial statements or in a separate GPFR;
- if the jurisdiction specific requirements specify that service performance information should be located in the same GPFR as the financial statements or in a separate GPFR.

Moreover, the *organization of service performance information* within a GPFR should enable users to understand an entity's service performance; to assess the entity's service efficiency and effectiveness and to use the service performance information for the purposes of accountability and decision-making.

14.5 A Comparison Between the Anvur Guidelines and RPG 3

The two documents represent a brief, not compulsory guide developed as a reference for the development of a PMS for the Italian and international context respectively. Analysis of the two documents aims to heighten understanding of whether the document developed by the Italian Agency is in line with the approach developed by IPSASB at international level.

Of course, the Italian guideline has been developed with the specific institutional and political Italian context in mind (Lapsley and Miller 2004); it does not make any explicit reference to international guidelines and it is mainly based on a rules-based standard setting. On the other hand, the RPG 3 follows the typical principles-based approach (Schaik 2010; Manes Rossi and Aversano 2015) of IPSASs that take into consideration the high diversity between the worldwide public sector entities.

The Italian guideline is not prescriptive but provides directions on the essential content considered by ANVUR to obtain a PMS for assessing the performance in relation both to the administration as a whole and the organizational units or areas of responsibility into which it is divided.

The main aim of the Italian guideline was the integration of the performance of administrative and teaching staff in order to create a continuous connection between them in all phases of the cycle performance. The integrated approach proposed by the ANVUR is a new approach because the original normative framework was mainly focused on the performance of the technical and administrative staff, while the teaching staff was regulated separately. This document however, evidences that a connection between the two areas, administrative and academic, increases the relevance of the performance in strategic decisions of universities.

The aim of the Italian guideline is consistent with the broader aims of the RPG 3 that represent good practice for developing a PMS that allows users to evaluate the services provided by public sector entities and to assess the entities' efficient and effective use of resources to deliver those services. Moreover, the RPG 3, in line with the Conceptual Framework of the IPSASB, emphasizes that the service performance information should be useful to users for accountability and decision-making purposes.

Moreover, the ANVUR guideline, in line with the RPG 3, evidences the importance of the so-called third mission, taking into consideration the impact of university activities on society and users involved; however, the guidelines do not specify the main category of users of the performance service information.

The ANVUR guideline gives more importance to the connection with the other Italian procedures of evaluation (VQR and AVA procedures) but does not give specifics on the connection between the performance measurement information and the financial statement. On the contrary, the IPSASB evidences the importance of this connection, including the possibility to present the performance measurement information in the same report as the financial statements.

In fact, RPG 3 allows entities to report service performance information either in the same report as the financial statements or in a separate report and, considering this option, the guideline indicates the information to disclose in both cases. This option has been included in order to give public sector entities the opportunity to observe their specific national jurisdictions. Therefore, it can be seen as a harmonization process that takes into account the specificities of individual countries.

Moreover, the RPG 3 evidences that the reporting service performance information should be annual, and it should cover the same period of reference as the financial reporting covers. This last indication is not explicated by the ANVUR guideline but it can be considered implicit. Moreover, the ANVUR, like the IPSASB, highlights that, in case of activities lasting more than a year, the university has to provide evidence of the phases of the activity and the expected interim results.

Both guidelines stress the importance of performance indicators; in particular, the ANVUR guideline evidences that the performance indicators supports the assessment of the cycle of the performance. Nevertheless, studies have proven that they should be given moderate importance to avoid falling into a performance paradox, i.e. the loss of connection between the performance measurement and the performance itself (Van Thiel and Leeuw 2002).

In terms of the types of indicators, the Italian guidelines specify that, in coherence with the specific objectives, specific indicators should be developed in relation to teaching, research and third mission related activities and indicators related to the administrative performance should be connected with them. This approach evidences the focus both on the single organizational units of the university (for example, the departments) and on the university as a whole.

IPSASB instead defines some types of performance indicators (inputs, outputs, outcomes, efficiency and effectiveness) but posits that service performance objectives may be expressed also using a narrative description of a desired future state. In

keeping with the principles-based approach of the IPSASB (Schaik 2010), these type of indicators can be applied to the variety of different services and different service delivery contexts that exist globally. However, studies show that the number of the indicators used should be moderate. A high number of them can undermine the comprehensibility and create an information overload, while a low number of indicators can facilitate the occurrence of a performance paradox (Van Thiel and Leeuw 2002).

Moreover, the indicators used should not always be the same to avoid incurring the “ossification” effect evidenced by Smith (1995); it can favour comparison but limit innovation.

The prevision of descriptive information, as an alternative or in combination with the quantitative indicators, is in line with the characteristics of a university, where some objectives cannot be measured by quantitative indicators (Cosenz 2011).

The characteristics required by the two guidelines are similar (pertinent, specific, measurable) and, taking into consideration the complexity of public sector entities, both of them evidence the possibility to develop indicators connected to transversal objectives (Broadbent and Laughlin 2009; Ferreira and Otley 2009).

Ultimately, IPSASB encourages the disclosure of all additional information useful for the users; similarly, the Italian guideline requires giving the information about the risk areas of corruption.

14.6 Conclusion

The Italian guidelines present a different structure from that provided by the IPSASB and follow a managerial approach with a focus on indicators, objectives and actors. Indeed, the IPSASB’s RPG 3 provides basic definitions and principles for the Presentation and Organization of Service Performance Information (Manes Rossi and Aversano 2015). The evolution of the Italian performance systems evidences that, after a long dominance of the bureaucratic model, under the NPM the reforming process has attempted to develop a managerial approach in the performance measurement of universities.

Nevertheless, the reform process, including the ANVUR guidelines, does not take into account the need for international comparison and the consequent harmonization problem. Even if a European harmonization process has been started with regard to the Quality Assurance in the European Higher Education Area (ESG), the ANVUR guideline does not make explicit reference to the IPSASB guideline for the performance measurement system. On the other hand, the analysis conducted evidences several similarities.

The Italian Agency, like the IPSASB, gives high importance to the performance indicators. Previous studies evidence the importance of both quantitative and qualitative performance indicators in assessing academic staff (Kallio and Kallio 2014); nevertheless, it seems that the Agency satisfies this balance only for the

research activity; the teaching activity does not use quantitative parameters. The subject of the evaluation are the departments, and the information produced is mainly addressed to the external and internal users for accountability and decision-making reasons; nevertheless, the external users have no power in the definition of the objectives and strategies of universities.

RPG 3 represents a point of reference for the harmonization of PMS and for the achievement of greater transparency and accountability. The gap evidenced by the present research shows that Italian universities are lagging behind in the harmonization of PSM, while evidencing that the institutional, cultural and historical contexts of each country can strongly influence the national approach to the development of PMS.

This research represents a preliminary documentary analysis; therefore, it has some limitations because it does not attempt to discuss the empirical implications of the national and international guidelines.

The research results would be beneficial both for public managers and policy-makers by highlighting how the Italian guidelines are not fully incorporating the approach followed on the international scene, reducing international comparability and, thus, accountability.

References

- Ammons, D., & Rivenbark, W. (2008). Factors influencing the use of performance data to improve municipal services: Evidence from the North Carolina Benchmarking Project. *Public Administration Review*, 68(2), 304–318.
- ANVUR. (2015). *Linee Guida per la gestione integrata del Ciclo della Performance delle università statali italiane*. Available at <http://www.anvur.org/attachments/article/806/Linee%20Guida%20Atenei.pdf>
- Arnaboldi, M., & Azzone, G. (2010). Constructing performance measurement in the public sector. *Critical Perspectives on Accounting*, 21(4), 266–282.
- Barrados, M., & Blain, J. (2013). Improving program results through the use of predictive operational performance indicators: A Canadian case study. *American Journal of Evaluation*, 34(1), 45–56.
- Behn, R. D. (2003). Why measure performance? Different purposes require different measures. *Public Administration Review*, 63(5), 586–606.
- Benito, B., Brusca, I., & Montesinos, V. (2007). The harmonization of government financial information systems: The role of IPSASs. *International Review of Administrative Sciences*, 73(2), 293–317.
- Beyle, H. C., & Parratt, S. D. (1938). Public attitudes and government efficiency. *Annals of the American Academy of Political and Social Science*, 199, 26–32.
- Bianchi, C., & Rivenbark, W. C. (2012). A comparative analysis of performance management systems: The cases of Sicily and North Carolina. *Public Performance & Management Review*, 35(3), 509–526.
- Bologna Working Group on Qualifications Frameworks. (2005). *A Framework for the Qualifications of the European Higher Education Area*. Copenhagen.
- Bouckaert, G., & Balk, W. (1991). Public productivity measurement: Diseases and cures. *Public Productivity & Management Review*, 15(2), 229–235.

- Broadbent, J., & Laughlin, R. (2009). Performance management systems: A conceptual model. *Management Accounting Research*, 20, 283–295.
- Brusca, I., Manes Rossi, F., & Aversano, N. (2015). Performance measurement in Italian and Spanish local governments: Comparative policy analysis. *Journal of Comparative Policy Analysis: Research and Practice*. doi:10.1080/13876988.2015.1094892.
- Cavalluzzo, K. S., & Ittner, C. D. (2004). Implementing performance measurement innovations: Evidence from government. *Accounting, Organizations and Society*, 29(3), 243–267.
- Christopher, J., & Leung, P. (2015). Tensions arising from imposing NPM in Australian public universities: A management perspective. *Financial Accountability & Management*, 31(2), 171–191.
- Cosenz, F. (2011). *Sistemi di governo e di valutazione della performance per l'azienda « Università »* (Vol. 51). Milano: Giuffrè Editore.
- Cuganesan, S., Guthrie, J., & Vranic, V. (2014). The riskiness of public sector performance measurement: A review and research agenda. *Financial Accountability & Management*, 30(3), 279–302.
- de Lancer, Julnes P., & Holzer, M. (2001). Promoting the utilization of performance measures in public organizations: An empirical study of factors affecting adoption and implementation. *Public Administration Review*, 61(6), 693–708.
- Esposito, V., De Nito, E., Pezzillo, Iacono M., & Silvestri, L. (2013). Dealing with knowledge in the Italian public universities: The role of performance management systems. *Journal of Intellectual Capital*, 14(3), 431–450.
- European Consortium for Accreditation (ECA). (2009). *Position paper on the BFUG—“Bologna beyond 2010” Report*.
- Ferreira, A., & Otley, D. (2009). The design and use of performance management systems: An extended framework for analysis. *Management Accounting Research*, 20(4), 263–282.
- GASB. (2010). *Suggested guidelines for voluntary reporting*. GASB: SEA Performance Information.
- Greiling, D. (2005). Performance measurement in the public sector: The German experience. *International Journal of Productivity and Performance Management*, 54(7), 551–567.
- Higgins, J. C. (1989). Performance measurement in universities. *European Journal of Operational Research*, 38(3), 358–368.
- Hood, C. (1995). The ‘New Public Management’ in the 1980s: Variations on a theme. *Accounting, Organizations and Society*, 20(2/3), 93–109.
- IPSASB. (2015). *Recommended Practice Guideline (RPG3) on Reporting Service Performance Information*. IFAC. Available at <http://www.ifac.org/system/files/publications/files/IPSASB-RPG-3-Reporting-Service-Performance-Information.pdf>.
- Johnsen, Å. (2005). What does 25 years of experience tell us about the state of performance measurement in public policy and management? *Public Money and Management*, 25(1), 9–17.
- Jones, C. S. (1991). Power, politics and the Jarratt proposals for accounting in British universities. *Financial Accountability and Management*, 7(3), 142–158.
- Kallio, K. M., & Kallio, T. J. (2014). Management-by-results and performance measurement in universities—Implications for work motivation. *Studies in Higher Education*, 39(4), 574–589.
- Kallio, K. M., Kallio, T. J., Tienari, J., & Hyvönen, T. (2016). Ethos at stake: Performance management and academic work in universities. *Human Relations*, 69(3), 685–709.
- King, Alexander F. (2000). The changing face of accountability: Monitoring and assessing institutional performance in higher education. *Journal of Higher Education*, 17(4), 411–431.
- Kloot, L., & Martin, J. (2000). Strategic performance management: A balanced approach to performance management issues in local government. *Management Accounting Research*, 11(2), 231–251.
- Kuah, C. T., & Wong, K. Y. (2011). Efficiency assessment of universities through data envelopment analysis. *Procedia Computer Science*, 3, 499–506.
- Lapsley, I. (2008). The NPM agenda: Back to the future. *Financial Accountability & Management*, 24(1), 77–96.

- Lapsley, I., & Miller, P. (2004). Foreword: Transforming universities: The uncertain, erratic path. *Financial Accountability & Management*, 20(2), 103–106.
- Leitner, K. H. (2004). Intellectual capital reporting for universities: Conceptual background and application for Austrian universities. *Research Evaluation*, 13(2), 129–140.
- Manes, Rossi F., & Aversano, N. (2015). Advancing performance measurement Italian local government vis-à-vis the IPSASB project. *The International Journal of Productivity and Performance Management*, 64, 76–93.
- Modell, S. (2003). Goals versus institutions: The development of performance measurement in the Swedish university sector. *Management Accounting Research*, 14(4), 333–359.
- OECD. (2007). *Towards Better Measurement of Government*, OECD Working Papers on Public Governance, 1, OECD Publishing. doi:10.1787/301575636734
- OECD. (2009). *Roadmap for the OECD Assessment of Higher Education Learning Outcomes Frasibility Study* (2nd version).
- Pettersen, I. J. (2014). A tale of diverse qualities—Reflections on performance measures in higher education. In A. Bourmistrov & O. Olson (Eds.), *Accounting, management control and institutional development* (pp. 201–216). Oslo: Cappelen Damm Akademisk.
- Poister, T. H. (2003). *Measuring performance in public and nonprofit organizations*. San Francisco: Jossey-Bass.
- Pollitt, C. (2013). The logics of performance management. *Evaluation*, 19(4), 346–363.
- Sanderson, I. (2001). Performance management, evaluation and learning in modern local government. *Public Administration*, 79(2), 297–313.
- Schaik, F. (2010). Materiality in government auditing. *International Journal on Governmental Financial Management*, 10(2), 2–12.
- Smith, P. (1995). On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration*, 18, 277–310.
- Sotirakou, T., & Zeppou, M. (2006). Utilizing performance measurement to modernize the Greek public sector. *Management Decision*, 44(9), 1277–1304.
- Sousa, C. A., de Nijs, W. F., & Hendriks, P. H. (2010). Secrets of the beehive: Performance management in university research organizations. *Human Relations*, 63(9), 1439–1460.
- ter Bogt, H. J., & Scapens, R. W. (2012). Performance management in universities: Effects of the transition to more quantitative measurement systems. *European Accounting Review*, 21(3), 451–497.
- Torres, L., Pina, V., & Yetano, A. (2011). Performance measurement in Spanish local governments. A cross-case comparison study. *Public administration*, 89(3), 1081–1109.
- van Dooren, W., & Van de Walle, S. G. J. (2008). *Performance information in the public sector: How it is used*. Basingstoke: Palgrave MacMillan.
- van Dooren, W., Bouckaert, G., & Halligan, J. (2015). *Performance management in the public sector*. New York: Routledge.
- van Helden, G. J., Johnsen, Å., & Vakkuri, J. (2012). The life-cycle approach to performance management: Implications for public management and evaluation. *Evaluation*, 18(2), 159–175.
- Van Thiel, S., & Leeuw, F. L. (2002). The performance paradox in the public sector. *Public Performance & Management Review*, 25(3), 267–281.
- Yang, K., & Holzer, M. (2006). The performance–trust link: Implications for performance measurement. *Public Administration Review*, 66(1), 114–126.

Chapter 15

Performance-Based Funding and Internal Resource Allocation: The Case of Italian Universities

Andrea Francesconi and Enrico Guarini

Abstract Management literature has emphasized the importance of performance measurement systems (PMS) that are consistent with the organization's context and strategy. However, little attention has been paid to factors that explain the design of PMS not adequately reflecting an organization's business model. We use the case of Italian universities to highlight how performance-based funding impacts resource allocation systems of faculty positions to academic departments. The findings show a variety of PMS in use. Differences across systems arise from different strategies and priorities. We have limited evidence that the systems' features can explain the disparities in research performance and funding, which suggest that part of the explanation may reside in the size of the organization or in the power of academic disciplines.

Keywords Performance measurement systems · Performance-based funding · Resource allocation · Organizational behaviour · Higher education

15.1 Introduction

The need for an efficient and effective performance management systems (PMS) within public institutions has increased over time. This is because it has been shown that the use of PMS improves the performance and overall quality of organizations (Lawson et al. 2003; De Bruijn 2002). Many public organizations have therefore introduced PMS in the belief that they will lead to better results and that they will increase organizational accountability both on an internal and an

A. Francesconi (✉)

Department of Economics and Management, University of Trento, Trento, Italy
e-mail: andrea.francesconi@unitn.it

E. Guarini

Department of Business Administration, Finance, Management and Law,
University of Milano-Bicocca, Milano, Italy
e-mail: enrico.guarini@unimib.it

© Springer International Publishing AG 2018

E. Borgonovi et al. (eds.), *Outcome-Based Performance Management in the Public Sector*, System Dynamics for Performance Management 2,
DOI 10.1007/978-3-319-57018-1_15

289

external side. The performance measurement (PM) literature highlights how PM systems should be integrated with strategic processes to enable organizations to achieve higher levels of performance and to satisfy stakeholders (Kaplan and Norton 1992; Neely 2005; Neely et al. 2005).

A central theme of approaches to new public management is the emulation of the market using state-induced competition. The allocation of state funds on the basis of comparative performances sets an incentive for competitive practices among universities. Policymakers, at the same time, have been actively seeking tools and instruments to get better performances from public organizations, especially when these organizations are largely funded by public finances. This has often occurred in the higher education sector in which one of the most popular approaches has been performance-based funding (PBF). Various forms of performance-based funding have been implemented in different countries (e.g., England, Scotland, Germany, Netherlands, Sweden, Denmark, the United States, and Chile), which have used both teaching/learning outcomes and research productivity as performance measures (Jongbloed 2001; Geuna and Martin 2003).

PBF uses a clearly specified formula to tie funding to organizational performances based on specific indicators and targets. The introduction of PBF may further impact organizations involved in higher education (i.e., Universities and Colleges) and affect the way in which they implement and/or modify their PMS to cope with PBF schemes (Burke and Serban 1997; Burke 2002).

Since Italian universities get their funding from public sources, our chapter is focused on two research questions:

- (a) How much are PM systems aligned with national performance-based funding?
- (b) Are universities' decision-makers aware of this alignment?

We expect that institutions with higher performance included in the PBF strongly value these indicators in their internal resource allocation system and performance measurement systems.

15.2 Literature Review

Two main different strands of literature are linked to our study:

- (a) On the one hand, performance measurement and performance management theories are applied both to private for profit and public not for profit entities.
- (b) On the other hand, performance-based funding studies and their implications in terms of public policy implementation are analyzed.

Referring to the first approach, we might see how performance measurement (PM) literature highlights how PM systems have to be integrated with strategic processes to enable companies and organizations to achieve higher levels of performance, to increase their outcomes and to satisfy all their stakeholders

(Kaplan and Norton 1992, 2001; Neely 2005; Neely et al. 2005). Furthermore, management control systems have been analyzed even in terms of their power to address organizational attention on strategic uncertainties, to guide the emergence of new strategies, and to ensure a continuing competitive advantage. Under this point of view, management control systems are used to focus organizational attention on strategic uncertainties, and thereby guide the development of new strategic initiatives (Burgelman 1983; Simons 1992, 2000). Contributors also highlight links between strategy and resource allocation systems (Bower 1970; Bower and Clark 2005). PM studies have focused their attention even on the relationships among control systems, resource sharing, and competitive strategies and their interactive effects on a business unit's performance (Govindarajan and Fisher 1990). These concepts have also been investigated by public management scholars (Osborne and Gaebler 1992; Osborne and Plastrik 2000; Bouckaert and Halligan 2008; Van Dooren et al. 2015). These scholars note various PM's strengths related to specific factors, such as transparency increases, incentives for better outputs, and incentives for shaping accountability (De Bruijn 2002). The use of formal performance measures based on explicit and objectively defined metrics has been a fundamental component of both public and private sector incentive systems. This literature has emphasized the importance of designing and using relevant measures that reflect the issues of relevance to the business (Lynch and Cross 1991), and the organization's context and strategy (Wagoner et al. 1999). However, the research indicates a large number of failures in performance measurement design (Pidd 2005). Furthermore, attention is paid to what factors explain the design of measurement systems that does not reflect adequately the organization's business environment (Dearden 1971, Paranjape et al. 2006).

Furthermore, scholars have noted some critical issues about the implementation of PM systems within public administration that highlight (Van Dooren et al. 2015) the dysfunctional behaviours of PMS, such as distortion of performance information (e.g., over and under representation, inflation of indicators, manipulation, and misinterpretation), and distortion of outputs that alter the daily functioning of the organization [i.e., measure fixation (Smith 1995; Moynihan 2009), cream skimming (Behn and Kant 1999), and myopia (Bouckaert and Balk 1991)]. These limits may be particularly relevant within professional organizations and, in particular, within sectors such as healthcare (Bevan and Hood 2006) and higher education, in which physicians and professors may encounter difficulties in accessing performance indicators (Jacobs et al. 2004). In these organizations, PMS may face serious problems in reflecting and adequately measuring 'good performance' of core processes, such as caring and research. Another important issue is decoupling performance indicators and goals that in some cases may lead to the non-use of PMS (Van Dooren et al. 2015). Decoupling, as opposed to the integration between goals and performance indicators, is a particular strategy used by organizations for seeking simultaneously legitimacy of different constituencies (Brignall and Modell 2000).

As for PBF, the literature highlights its impact within the field of higher education in which many PBF projects have been implemented in various countries all over the world. PBF first appeared in the late 1970s but was abandoned later. However, in

recent times, it has been reintroduced by many states within the USA and by other relevant European countries, such as the United Kingdom and Italy (Harnisch 2011; Jones 2012). Essential research questions related to PBF are as follows:

- (a) The impact on organizational outcomes (Layzel 1999);
- (b) The way these impacts are produced (Massy 1996);
- (c) The obstacles and the unintended effects encountered in the process of introducing PBF (Liefner 2003).

The research indicates links and connections between PBF programmes and specific theories of action (Argyris and Schon 1986) that translate substantive policy goals into concrete actions (McDonnell and Elmore 1987). The most cited theory of action is that institutional performance will be improved through material incentive that mimics the profit motive for business organizations (Burke 2005). This material incentives theory of action holds that higher education organizations will make strong efforts to improve their performance if the amount of funding involved is significant enough.

Advocates of performance-funding programmes have exposed other theories of action, with the following specific aims:

- (a) To provide information about the goals as a mean to catalyze institutional change and persuade the faculty. To have effective resource allocation, a good understanding of the proper balance between intrinsic values of the higher education institutions and their environment is needed;
- (b) To make institutions aware of their performance and to implement benchmarking processes in order to mobilize feelings of pride and status striving. One PBF-based formula with major advantages is the use of objective criteria, which provides clear insight regarding the distribution of funds among universities and, therefore, facilitates comparisons between institutions and reduces the lobbying by institutions (Burke 2005; Dougherty and Hong 2006).
- (c) To increase organizational awareness of gaps in their performance relative to their own goals and standards and to promote experiential learning through organizational self-appraisal (Huber 1991).

PBF theories very often consider and highlight the changes and efforts needed at the organizational level (i.e., changes in support services for research and teaching) to reach the higher level of outcomes pursued. A special focus is on the need to manage complexity at the organizational level. Universities are professional organizations in which the expertise needed for effective teaching and effective research lies almost exclusively in the hands of the faculty (Massy 1996).

Research has considered even the perils and risks of PBF with special attention on (Dougherty and Reddy 2013)

- (a) The unintended side effects, such as the weakening of academic standards and excessive focus posed on financed outcomes. The forms of resource allocation influence the behaviour of academics and managers in higher education, particularly their level of activity as well as the kinds of activities they engage in

and their ways of dealing with risks (Liefner 2003). Burke (1997) also laments the lack of equity and access indicators in many PBF programmes. Additionally, PBF presents practical problems. Assigning funding weights for indicators involves decisions on whether to weigh all measures equally or to assign distinct values to each item. Funding levels have to find that delicate balance between amounts large enough to spur improved performance yet not so large as to threaten budget instability. PBF reflects with varying emphasis the policy values of efficiency, quality, equity and choice, but the past, PBF programmes stressed efficiency and quality rather than equity and choice indicators, which were slighted (Burke and Modarresi 2000).

- (b) Some technical problems have occurred. For example, the formula may prove inadequate in situations where changing needs and client bases are to be tackled. Additionally, the formula may be inadequate for dealing with differentiation among units and institutions (Massy 1996).
- (c) The obstacles to overcome include the poor definition of performance indicators, inequality in institutional capacity to diagnose performance problems, and difficulties in finding out workable solutions (Massy 1996; Burke 2002).

These perils may, at the end, lead to a common level of mediocrity. In fact, each unit (i.e., university or department) is funded on the same quantitative grounds. In this context, the use of formulas may reduce the incentives to seek outside funds and perpetuate funding inequities because the units tend to become fixated on the parameters driving the formula. In our study, we adopt the basic assumption of the contingency approach to investigate how individual universities adapt their PMS to PBF mechanisms. The contingency approach emphasizes the importance of the situational influence on the management of organizations and may be useful to understand the various factors that influence the use of PMS in universities. The essential premise of the contingency approach (Tosi and Slocum 1984) is that organizational effectiveness, broadly defined as adaptation and survival, can be achieved using a variety of paths and ways. Each way is not effective under all conditions; certain organizational actions or responses are more appropriate than others, depending on the situation (Galbraith 1977). Referring to our specific research, the contingency approach affirms that there is no universally appropriate PM system, but that particular features of PMS and their effectiveness will depend on specific organizational and contextual factors (Otley 1980), including organizational structure (Ferreira and Otley 2010), strategy (Maltz et al. 2003), size and technology (Chenhall 2007), culture and leadership (Moynihan and Ingraham 2004; Berry et al. 2009).

15.3 The Context of PBF in Italian Public Universities

The Italian Higher Education system is comprised of 67 public universities, and 30 private universities.

Universities have organizational autonomy within national administrative and funding rules. Public universities are mainly financed through central government transfers, which currently accounts for approximately 80% of their revenue with the remaining 20% coming from private funds (i.e., mainly tuition fees). Tuition fees of universities are capped at 20% of the main central government funding for financing universities' operations, which is known as "Fondo di Finanziamento Ordinario" (FFO). Central transfers are provided for the construction of university buildings, and development programmes.

In the past, the FFO fund has been allocated to public universities mainly on an expenditure basis. In 2014, the Ministry of Higher Education introduced a performance-based funding system, which was based on research and teaching performance indicators. The system is formula-based with an annually increasing weight of performance-based indicators. Currently, the FFO fund is allocated for 20% on a performance basis (i.e., 18% in 2014–2015) and for 80% on an expenditure basis. In particular, the latter is allocated for 20% on a standard cost per regular student (i.e., increasing up to 100% in 2018), and the remaining 80% is allocated on the previous year's expenditure allocation. The 20% performance-based share of FFO is allocated as follows:

1. 70% based on university performance in the national research assessment framework "Valutazione della Qualità della Ricerca" (VQR);
2. 10% based on the quality of recruitment (i.e., the VQR research performance of newly hired faculty);
3. 20% based on the internationalization of teaching activities (i.e., number of Erasmus students and credits from courses attended overseas).

It is worth noting that the share of transfers allocated on the standard cost per student is, essentially, a performance-related funding system. Indeed, each regular student is recognized with a certain amount of direct cost (i.e., the cost of teaching and operations), depending on the type of academic programme (i.e., social sciences, health sciences, and technology). The cost per student is calculated yearly for each university, depending on the real costs compared to the standard, and then the overall funding is distributed according to the share of each university. Hence, the universities showing real costs below the average and the number of regular students aligned with programme standards are recognized according to "higher standard cost per student" and receive proportionally higher transfers. This approach focuses on input and output measures as the formula provides incentives for cost efficiency and the increase of students' productivity (i.e., regular students). An important component of the resource allocation system of Italian universities is the budget authorization for faculty hiring released by the central government. The system is based on the allocation of "hiring points", which consists of full-time equivalents of faculty and staff members that universities are allowed to hire in the next fiscal year. The yearly hiring points for the overall system are allocated to individual universities on the basis of their financial performance, which is calculated on the basis of the economic margin, i.e., the difference between total

revenues (performance-based funding, standard cost transfers, other central transfers and tuition fees) and the net cost of faculty and staff. Hence, universities with higher economic margins get higher shares of overall hiring points. It should be noted that the allocation of hiring points is interlinked with student-related funding since both standard cost transfers and tuition fees are included in the calculation of total revenues.

15.4 Data and Methods

This study is focused on the Italian Higher Education System. At the central level, the Ministry for Higher Education and Research has developed a PBF system to improve university performance. The PBF's aim is to allocate a share of the national budget on the basis of performance indicators. For each university, the internal resource allocation system and its performance indicators basis is broadly considered in this study as performance measurement system (PMS).

In particular, we have considered the internal systems used by universities for allocation of hiring points to departments as this is one of the most important decisions made by a university, which is affected in Italy by a centrally led performance-based resource allocation system.

We used a case study approach because it is an appropriate methodology to investigate a complex phenomenon that has many variables (Yin 1994). The data were gathered from different sources to ensure triangulation and the internal validity of the empirical evidence (Stake 1995). In particular, we have considered data from the Italian Ministry of Higher Education, public data from university websites, and the university's internal data.

First, we have analyzed the primary documents that are publicly available to collect information regarding the features of internal resource allocation system. The universities have been classified according to the type of resource allocation system. Hence, we have identified two main groups of universities, those using a formula-based resource allocation system, and those not using a formula-based system. Among the latter, we distinguished those cases where decision-making is supported by the use of quantitative indicators, but allocation is not based on formulas and the university's top management uses a certain amount of discretionary assessment. Moreover, we considered the ranking of the university in the last 2010–2014 VQR performance, which has been considered by the central government as the basis for the allocation of performance-based transfers since FY 2014.

In the case of large universities, we attributed the value “High” to universities ranked from 1st to 10th place in the VQR ranking; value “Medium” to universities ranked from 11st to 20th place; and value “Low” to universities ranked over the 20th place.

In the case of medium-sized universities we attributed the value “High” to universities ranked from 1st to 8th place in the VQR ranking; the value “Medium”

to universities ranked from 9th to 16th place' and the value "Low" to universities ranked over the 16th place.

Second, we analyzed the share of VQR performance-based funding on university total revenues. We observed a normal distribution of individual shares ranging from 6.81 to 15.38%.

Third, we selected seven relevant cases with best (i.e., high-ranked) and worse (i.e., low-ranked) VQR performance and with minimum, medium and maximum shares of performance-based transfers on total revenues. Among the seven universities, four were selected from the top five performers (i.e., two from large universities and two from medium-sized universities), and the three others from the five worse performers (i.e., two from large universities and one from the medium-sized universities).¹ The size of universities considered in the research ranking is based on the number of faculty members assessed. Whereas the data related to internal resource allocation systems for selected relevant cases were not publicly available, a specific request was addressed to these universities. Here, two top performers among large universities and one from the medium-sized universities denied access to internal PMS features.

This sample is consistent with the aim of theoretical sampling and theory building (Eisenhardt 1989) given that the purpose of the study is to increase knowledge on a relatively unexplored topic.

We interviewed the rector of each university since they are the head of the university's governing body, making decisions on internal resource allocation (i.e., in University D, we interviewed the vice-rector).

The interviews were conducted jointly by the two authors of this study, and lasted for approximately 30–45 min each. Each interview was recorded and transcribed. Interviews were focused around the following questions: (a) the motivation for adopting a performance-based resource allocation and their alignment with the PBF system; (b) the organizational process adopted for the introduction of the system; and (c) the impact within the organization. Once collected, the data were analyzed following a two-step protocol.

First, we reordered and categorized the interview data according to the features of the PMS. We distinguished between the data extracted from the university financial reports, and the statements of the interviewees. Second, in the interpretation step, we examined the features of the PMS and their alignment with the university's strategy and funding.

Each step has been conducted independently by each of the investigators of this study, and the findings were compared and interpreted jointly.

Table 15.1 shows some structural characteristics of the selected universities, such as the total amount of students, the faculty size, the revenues, and the VQR ranking.

The data show a large variability in terms of total revenue (i.e., coherent with the different dimensions of these universities) as well as significant differences in terms

¹To maintain anonymity, the university names have been replaced by letters.

Table 15.1 Main characteristics of selected universities

University	Total students (academic year 2014/2015)	Faculty (FTEs)	Total revenues (Euro millions)	Position in the in Italian VQR ranking
A	31,593	883	140 < X < 180	High
B	15,515	614	80 < X < 120	High
C	10,367	383	30 < X < 70	High
D	41,840	1554	200 < X < 250	Low
E	17,909	694	80 < X < 120	High
F	21,165	921	140 < X < 180	Low
G	46,748	1301	200 < X < 250	Low

Source Data from www.miur.it

of total students. Table 15.2 shows the main features of the internal resource allocation systems based on departments' performance.

15.5 Findings and Discussion

The findings emerging from the analysis of our sample allows us to note some preliminary answers to the initial research questions.

First, it is possible to highlight consistency between the PMS of the universities and their strategies. This consistency is stronger whether the strategies are highly formalized, and evident whether the strategy has been explicitly remarked. In other cases, it is not possible to find any kind of relationship.

In the cases analyzed, interviews with the rectors also highlight a substantial consistency of PMS with the funding model of the Italian University System. This evidence is confirmed by the high degree of awareness about this consistency at the key decision-maker level.

The data show an effective design of PMS in terms of their alignment to PBF indicators, which confirms that PBF is influencing universities' decision-making and managerial tools (Burke 2002).

Only in the case of University A has the rector highlighted a strong orientation of the system towards the research dimension of performance, which is the major strategic goal of that institution:

We are a research University [...] All our teaching activities are research-driven... We cannot allocate new hiring points to departments on the basis of teaching indicators, such as the number of students enrolled in our programmes because we do not simply teach.

In this case, moreover, we find a low degree of awareness regarding the weight of student-related funding and the real functioning of the financing system:

Table 15.2 Features of resource allocation systems in the selected sample

University	Student-related funding/total revenues (%)	VQR-research performance funding/total revenues (%)	Allocation basis	Performance indicators	Algorithm (weights)
A	43	12	(a) Research (b) Teaching (c) Faculty retirements	(a) VQR-Research performance (b1) Regular students (b2) Amount of credits (c) Number of retirements in the previous three years	(a) 50% (b) 30% (b1 15% and b2 15%) (c) 20%
B	34	14	Non-formula based	–	–
C	34	14	Non-formula based	–	–
D	26	8	Two criteria: (a) Related to PBF (b) Not related to PBF (i.e., internal sustainability)	(a1) VQR-Research performance (a2) Teaching (a3) Rate of internationalization (i.e., students) (b1) Faculty members retirement (b2) Number of habilitated faculty for professorship (b3) Faculty (i.e., total number)	(a1) 24% (a2) 70% (a3) 6% (b1) 45% (b2) 30% (b3) 25%

(continued)

Table 15.2 (continued)

University	Student-related funding/total revenues (%)	VQR-research performance funding/total revenues (%)	Allocation basis	Performance indicators	Algorithm (weights)
E	28	9	Non-formula based (i.e., use of indicators)	(a) Teaching attractiveness and quality (b) VQR-Research performance	No algorithms: (a) 50% (b) 50%
F	35	11	Three criteria: (a) Strategic priorities (b) Funded national plan for associate professorship hiring (c) Related to PBF	(a) Rector decision (b) Habilitated faculty for professorship (c) Teaching, research and fund raising	(a) 14% (b) 21.85% (c) 64.15%
G	31	8	Different criteria for different kinds of allocation decisions: Full Professor: (a) Quality of research (b) Retirements Associate Professors: (a) Teaching needs (b) Research performance (c) Retirements (d) Number of habilitated professors Assistant Professors: (a) Research performance (b) Equity between disciplines	(a1) VQR-research performance (a2) Habilitated faculty for Professorship retirements (b) Number of retirements (c) Weighted students per faculty (d) VQR-Research (e) % of retirements (g1) VQR-research performance (h2) per unit allocation	(a) and (b) allocated proportionally, no algorithm (c) 50% (d) 30% (e) 20% (g1) 33.3% (h2) 66.6%

Source Authors' extract from university documents

The PM system adopted by our university is perfectly aligned with the national performance-funding system. [...] In fact, our internal criteria reflect perfectly the national ones.

These sentences are not coherent with the financial data since, as already seen in previous sections of this chapter, the average weight of research performance-related funds is very low compared to the overall budget of Italian universities and to the overall funding scheme.

In all other cases, the evidence from the interviews shows how universities are trying to link their PMS with strategies. University D, for instance, has strongly considered the teaching perspective in designing its PMS:

Our University is facing a higher rate of student failure, and improving this issue is one of the most important goals of our strategic plan. Although in recent years, we faced a change in the governance of the University, there was no discontinuity of strategy; indeed, there is a strengthened attention to all issues related to the teaching process.

University F's PMS also appears to be strictly linked to the strategy as the rector pointed out:

Performance indicators and their weights have been identified in order to tie them with our strategy. We tried to keep coherence between strategic goals related to teaching and research activities and the indicators used for allocating faculty hiring points to departments.

In two out of the seven cases (i.e., B and C), the universities have not implemented formula-based PMS, which is the case of two of the top ranked medium-sized institutions in the Italian VQR. These two examples show a non-use of (i.e., formalized) PMS, which seems to confirm previous literature propositions (Moynihan 2008; Van Dooren et al. 2015). However, here, it seems that the steering and control function, which is one of the main uses of PMS in organizations (Van Dooren et al. 2015), is connected to different logical and organizational mechanisms. In these cases, "clan control" systems (Ouchi 1979) and quality appear to be particularly strong for resource allocation of faculty members. The top leaders consider these mechanisms to be the key drivers for managing the university and for reaching performance excellence, which seems to foster that PMS based on quantitative indicators and algorithms are not always well suited for a professional context such as universities.

The rector of University B told us:

PMS are useful, but we must not become their slaves. Indicators and algorithms are a compass you must direct, but then again, if one has clear what must be the kind of people needed then [...]. We search and strive for quality people [...] we use the compass of quality and we hire best performing people in the perspectives we consider relevant to our functioning: research, teaching and fund raising.

On the same topic, the rector of University C noted:

We got excellent results and high ranking just because our faculty did well! In our university, all faculty members have a high research potential, they publish successfully, they all teach a lot, they do well [...] Results can't be worse if you have hired excellent people

[...] I have not the problem of “measuring”. My major issue is that we can’t reward enough our faculty [...]. Anyway, I am aware of being managing a small organization and that in the case of large ones, PMS may be useful.

The introduction of performance-based funding has had an impact on the design of PMS of Italian universities. Italian universities have adopted different behaviours and models for implementing their PMS multidimensional systems. The large majority of Italian universities have opted for multidimensional formula-based PMS, while others did not use formulas for internal allocation purposes. Formula-based PMS are generally designed to be coherent with the PBF system. It is worth noting some relevant exceptions to this evidence.

Indeed, some universities have taken into account the PBF perspectives and indicators while they have decided to customize their PMS by introducing performance perspectives and indicators consistently with their strategies, budget constraints, and organizational structure.

One university (E) has taken a proactive stance by introducing in its PMS some of the indicators expected to rise in future PBF transfers (e.g., the standard cost per student).

Differences in the implementation strategies of PMS are evident even with regard to the choices of algorithms and indicators. In many cases, as emerging from the interviews (i.e., B, D and E), universities shy away from a mechanical and uncritical use of quantitative allocation criteria that can potentially lead to paradoxes and that cannot be aligned with its strategy. In other cases, algorithms defined by external parties, as well as PBF indicators, are seen as important drivers for introducing behavioural changes within Italian universities. For example, the rector of University D noted:

[...] in the end, we tried to look for a certain degree of confidence with PBFS. For example, the indicator of regular students (that we called “sustainability” in our internal system) was introduced for the first time in our university on a “standard cost basis” which differs a lot from the indicator used in PBF. Such system was certainly a novelty, but at the end it counts for 49% of the total amount of our formula-based system, almost the same weight the standard cost funding has within the PBF. It is not correct to say that there was a complete alignment with the weights of indicators in the national performance-based funding [...]. In numerical terms there were of course some differences in our formula-based system, but we tried to cope with the PBF with regards to the weight of various indicators [...]. We made some adjustments because we wanted to give significant relevance to standard cost funding and not only to VQR- performance funding.

The view of the rector of University G is also interesting:

When an organization begins to apply criteria and formula-based systems it also begins to be aware of their limits. Quantitative criteria and algorithms defined by external bodies are also useful in order to limit opportunistic behaviours and to avoid internal “criteria skimming”. But you have to consider that there is not a best and unique way to apply these criteria! [...]. You cannot manage a complex organization like if you were driving a tank in the garden! Our strategy for designing and implementing the next PMS will be to rely on criteria and parameters able to cope with our organizational needs and performance excellence.

15.6 Conclusions

This chapter aimed to contribute to performance measurement and performance evaluation studies in an expanding and little explored field of research, university management.

Our study aimed at highlighting in the higher education sector a classical research topic in managerial accounting: the relevance of performance management systems, their coherence with decision-making processes and strategies under a contingency approach.

Whether the PMS of universities are adopted for resource allocation, they show the following features:

- (A) They are multidimensional and balanced. Different performance perspectives are taken into consideration and their weights seem to be balanced.
- (B) They are focused on addressing faculty and organizational behaviours based on the achievement of strategic goals.

The PMS of the universities are characterized by a high degree of formalization (Amigoni 1992), which seem to be coherent with the universities' strategies. This result confirms, once again, a very consolidated theoretical issue (Kaplan and Norton 1992). Of course, when strategy has not been clearly remarked by interviewees (i.e., three cases out of seven), we have not been able to check for coherence.

PMS are largely coherent with organizations' structures and mechanisms, and this finding is once again coherent with the performance measurement literature.

In general, it can be concluded that performance-based resource allocation systems are used extensively at the university level and that their design is broadly influenced by PBF. However, our study provides evidence that PBF's influence is partial in that the relevance and weight of performance indicators used by universities are not strictly linked to the ones used in the national performance-funding scheme. These findings, contrary to Burke (2002), seem to indicate a high degree of maturity in the use of PMS by Italian universities, thereby reducing the risk of strictly focusing just on PBF measures (Jongbloed 2001). Case A, however, shows that some PBF pitfalls can occur when ignoring the real financial impact of research performance on the university budget. Indeed, research performance-based funding determines only a marginal part of total budget allocations (see again section 15.3).

One possible explanation emerging from case A may be that resource allocation is driven mainly by research performance whose relevance for funding is not equally weighted between the PBF and the university's business model (i.e., the relative weight of revenue sources).

In any case, the internal resource allocation systems of universities are very differentiated in terms of focus, measures, and use or formulas.

Differences across systems arise from differences in strategies and priorities of universities. In small universities, there is a lower use of formulas based on the awareness that PMS are not very helpful in this case since other managerial tools can drive better decision-making and faculty performance. Conversely, in the case of large universities, PMS play a major role in limiting discretion and providing transparency in decision-making (Jongboed 2001).

Nevertheless, we have limited evidence that the systems' features can explain the disparities in research performance and funding across universities. The findings show that good research performance can be achieved even without formula-based PMS, both in terms of small- and large-sized universities. This issue seems to suggest that driving faculty towards research excellence is not necessarily linked to the exclusive use of quantitative PMS, but to other management systems and values (i.e., quality of hiring, faculty values and ethics).

This study highlights that PMS can be used effectively for driving the organization towards strategic goals even in a professional context. At the same time, our findings show that, in a certain context, cultural biases might lead university leaders to emphasize the consistency of the internal formula-based resource allocation system with PBF, while ignoring the university's business model. However, this study does not provide enough evidence of the determinants of such cultural bias. Further studies should investigate this issue in greater detail.

Finally, we are aware that this work is designed as a case study, and we wonder about the generalizability of our results due to the small sample. We expect that by enlarging the sample in a longitudinal perspective and considering different national contexts, future studies might be useful for conducting an investigation if the use of PMS is more concerned with the exercise of power rather than with steering the organization.

References

- Amigoni, F. (1992). Planning management control systems. In C. Emmanuel, D. Otley, & K. Merchant (Eds.), *Readings in accounting for management control, part one* (pp. 174–185). London: Chapman & Hall.
- Argyris, C., & Schon, D. A. (1996). *Organizational learning II: Theory and practice*. Reading, MA: Addison Wesley.
- Behn, R. D., & Kant, P. A. (1999). Strategies for avoiding the pitfalls of performance contracting. *Public Productivity and Management Review*, 22, 470–489.
- Berry, A., Coad, A., Harris, E., Otley, D., & Stringer, C. (2009). Emerging themes in management control: A review of recent literature. *The British Accounting Review*, 41(1), 2–20.
- Bevan, G., & Hood, C. (2006). What's measured is what matters: Targets and gaming in the British health care sector. *Public Administration*, 84, 517–538.
- Bouckaert, G., & Balk, W. (1991). Public productivity measurement; diseases and cures. *Public Productivity and Management Review*, 15, 229–235.
- Bouckaert, G., & Halligan, J. (2008). *Managing performance: International comparisons*. London: Routledge.

- Bower, J. L. (1970). *Managing the resource allocation process: A study of corporate planning and investment*. Boston, MA: Harvard Business School, Division of Research.
- Bower, J. L., & Gilbert, C. (2005). *From resource allocation to strategy*. Oxford: Oxford University Press.
- Brignall, S., & Modell, S. (2000). An institutional perspective on performance measurement and management in the "New Public Sector". *Management Accounting Research, 11*, 281–306.
- Burgelman, R. A. (1983). Corporate entrepreneurship and strategic management: Insights from a process study. *Management Science, 29*(12), 1349–1364.
- Burke, J. C. (2002). *Funding public colleges and universities for performance: Popularity, problems, and prospects*. Albany, NY: Rockefeller Institute Press.
- Burke, J. C. (2005). The three corners of the accountability triangle: Serving all submitting to none. In J. C. Burke (Ed.), *Achieving accountability in higher education* (pp. 296–324). San Francisco, CA: Jossey Bass.
- Burke, J. C., & Modarresi, S. (2000). To keep or not to keep performance funding: Signals from stakeholders. *The Journal of Higher Education, 71*(4), 432–453.
- Burke, J. C., & Serban, A. M. (1997). *Performance funding and budgeting for public higher education: Current status and future prospects*. Albany, NY: Rockefeller Institute of Government.
- Chenhall, R. (2007). Theorising contingencies in management control research. In C. S. Chapman, A. Hopwood, & M. D. Shields (Eds.), *Handbook of management accounting research* (pp. 163–206). Oxford: Elsevier.
- De Bruijn, H. (2002). Performance measurement in the public sector: Strategies to cope with the risks of performance measurement. *The International Journal of Public Sector Management, 15*(7), 578–594.
- Dearden, J. (1971). *What's wrong with your financial control system* (pp. 27–35). Paris: European Business.
- Dougherty, K. J., & Reddy, V. (2013). *Performance funding for higher education: What are the mechanisms, what are the impacts?*. San Francisco, CA: Jossey-Bass.
- Dougherty, K.J., & Hong, E. (2006). Performance accountability as imperfect panacea: The community college experience. In T. Bailey & V.S. Morest (Eds.), *Defending the community college equity agenda* (pp. 51–86). Baltimore, MD: John Hopkins University Press.
- Eisenhardt, K. M. (1989). Building theories from case study research. *The Academy of Management Review, 14*(4), 532–550.
- Ferreira, A., Otley, D. (2010). Design and use of management control systems: An analysis of the interaction between design misfit and intensity of use, Working paper, Chartered Institute of Public Finance and Accountancy, London: CIPFA.
- Galbraith, J. R. (1977). *Organisational design*. Reading, MA: Addison-Wesley.
- Govindarajan, V., & Fisher, J. (1990). Strategy, control systems, and resource sharing: Effects on business-unit performance. *Academy of Management Journal, 33*, 259–285.
- Geuna, A., & Martin, B. R. (2003). University research evaluation and funding: An international comparison. *Minerva, 41*(4), 277–304.
- Harnisch, T. (2011). *Performance based funding: A re-emerging strategy in public higher education financing*. Higher Education Policy Brief. Washington, DC: American Association of State Colleges.
- Huber, G. F. (1991). Organizational learning: The contributing process and the literatures. *Organization Science, 2*(1), 88–115.
- Jacobs, K., Marcon, G., & Witt, D. (2004). Cost and performance information for doctors: An international comparison. *Management Accounting Research, 15*(3), 337–354.
- Jones, D. P. (2012). *Performance funding: From idea to action*. Washington, DC: Complete College America.

- Jongbloed, B. (2001). Performance based funding in higher education: an international review. Working paper no. 35, Monash University, ACER, Centre for the Economics of Education and Training.
- Kaplan, R. S., & Norton, D. P. (1992). The balanced scorecard: Measures that drive Performance. *Harvard Business Review*, 58(1), 686–705.
- Kaplan, R. S., & Norton, D. P. (2001). The strategy-focused organization. *Strategy and Leadership*, 29(3), 41–42.
- Lawson, R., Stratton, W., & Hatch, T. (2003). The benefits of a scorecard system. *CMA Management*, 77(4), 24–26.
- Layzel, D. T. (1999). Linking performance to funding outcomes at the state level for public institutions of higher education: Past, present, and future. *Research in Higher Education*, 40(2), 233–246.
- Liefner, I. (2003). Funding, resource allocation and performance in higher education systems. *Higher Education*, 46(4), 469–489.
- Lynch, R. L., & Cross, K. F. (1991). *Measure up!: Yardsticks for continuous improvement*. Cambridge, MA: Blackwell.
- Maltz, A. C., Shenhar, A. J., & Reily, R. R. (2003). Beyond the balance scorecard: Refining the search for organizational success measures. *Long Range Planning*, 36(2), 187–204.
- Massy, W. F. (Ed.). (1996). *Resource allocation in higher education*. Ann Arbor, MI: University of Michigan Press.
- McDonnel, L. M., & Elmore, R. F. (1987). Getting the job done: Alternative policy instruments. *Educational Evaluation and Policy Analysis*, 9(2), 133–152.
- Moynihan, D. P. (2009). Through a glass, darkly. Understanding the effects of performance regimes. *Public Performance & Management Review*, 32(4), 592–603.
- Moynihan, D.P. (2008). *The Dynamics of Performance Management: Constructing Information and Reform*. Washington DC: Georgetown University Press.
- Moynihan, D., & Ingraham, P. (2004). Integrative leadership in the public sector: A model of performance information use. *Administration and Society*, 36(4), 427–453.
- Neely, A. (2005). The evolution of performance measurement research: Developments in the last decade and a research agenda for the next. *International Journal of Operations & Production Management*, 25(12), 1264–1277.
- Neely, A., Gregory, M., & Platts, K. (2005). Performance measurement system design: A literature review and research agenda. *International Journal of Operations & Production Management*, 25(12), 1228–1263.
- Otley, D. (1980). The contingency theory of management accounting: Achievement and prognosis. *Accounting, Organizations and Society*, 5(4), 413–428.
- Osborne, D., & Gaebler, T. (1992). *Reinventing government*. New York: Penguin Books.
- Osborne, D., & Plastrik, P. (2000). *The reinventors filebook: Tools for transforming your government*. San Francisco, CA: Jossey-Bass.
- Ouchi, W. G. (1979). A conceptual framework for the design of organisational control mechanisms. *Management Science*, 25(9), 833–848.
- Paranjape, B., Rossiter, M., & Pantano, V. (2006). Performance measurement systems: Successes, failures and future—a review. *Measuring Business Excellence*, 10(3), 4–14.
- Pidd, M. (2005). Perversity in public service performance measurement. *International Journal of Productivity and Performance Management*, 54(5/6), 482–493.
- Simons, R. (1992). The role of management control systems in creating competitive advantage: New perspectives. In C. Emmanuel, D. Otley, & K. Merchant (Eds.), *Readings in accounting for management control* (pp. 622–645). London: International Thomson Business Press.
- Simons, R. (2000). *Performance measurement and control systems for implementing strategy*. Englewood Cliffs, NJ: Prentice Hall.

- Smith, P. (1995). On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration*, 18, 277–310.
- Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.
- Tosi, H. R., Jr., & Slocum, J. W. (1984). Contingency theory: Some suggested directions. *Journal of Management*, 10, 9–26.
- Van Dooren, W., Bouckaert, G., & Halligan, J. (2015). *Performance management in the public sector*. London: Routledge.
- Wagoner, B., Neely, A. D., & Kennerley, M. P. (1999). The forces that shape organizational performance measurement systems: An interdisciplinary review. *International Journal of Production Economics*, 60–61(3), 53–60.
- Yin, R. K. (1994). *Case study research: Design and methods*. Thousand Oaks, CA: Sage.

Chapter 16

Performance Reporting in Italian Public Universities: Activities in Support of Research, Teaching and the “Third Mission”

Elisa Bonollo and Mara Zuccardi Merli

Abstract Italian public universities are now paying close attention to performance measurement under the new public management process. This chapter aims to verify whether and how Italian public universities report the performance of activities in support of research, teaching and the “third mission”. Such can strongly impact how the university stakeholders will perceive the results of institutional operations. A documentary analysis of the Italian universities’ annual performance reports, for the years 2011–2014, was adopted to achieve this. The results show that further improvements can be made, especially in terms of outcomes, effectiveness, efficiency and coordination as far as the financial statements are concerned.

Keywords Reporting · Performance · Indicators · Public universities · Support services

16.1 Introduction

Over the past few decades, many studies have been conducted to identify the purpose and meaning of performance measurement in public organizations: focusing on both the international scenario (Behn 2003; Bouckaert and Hallingan 2008; Johnsen 2005) and especially on the public organizations of specific countries, including Italy (Anselmi 2009; Borgonovi 2002; Mussari 2001).

Italian public universities have devoted a great deal of attention to performance measurement in recent years. This trend has emerged following legislative reforms that affected all public organizations (Law 15/2009 and Legislative Decree 150/2009), especially public universities (Law 240/2010 and enabling acts), which

E. Bonollo (✉) · M. Zuccardi Merli
Department of Economics and Business Studies, University of Genoa, Genoa, Italy
e-mail: bonollo@economia.unige.it

M. Zuccardi Merli
e-mail: zuccardi@economia.unige.it

strive to address increasing demands in terms of accountability and transparency and to compete at an international level as far as the research, teaching and the “third mission” are concerned (Holzer and Halachmi 1996).

This chapter focuses on the performance reporting concerning the activities in support of research, teaching and third mission that Italian public universities must now implement and make public. Such vast services are carried out by these universities nonteaching staff, which greatly influences both the results of said core institutional activities and how the stakeholders perceive the quality of the same (Rebora 2003). Thus far, however, literature has generally overlooked the performance measurements with regards to universities. Therefore, it is now important to understand whether and how the performance of the abovementioned supporting services is measured, since accurate performance measurement is the first step needed to improve the effectiveness and efficiency of university management (Bianchi 2016; Otley 1999).

In this context, the chapter aims:

- To verify if the Italian public universities have prepared and published performance reports on their websites in accordance with the managerial principles;
- To identify what (which type of services) and how (by what type of performance indicators) public universities actually measure and report their performance in such documents.

A documentary analysis of the performance reports of all Italian public universities, for the years 2011–2014, was adopted to achieve this and using a spatial comparison to highlight the information of their content.

The remainder of the chapter is structured as follows: the next section covers past literature on performance reporting within public organizations in general, as well as the Italian legislative initiatives on public performance measurement. The third section focuses on the proposal of scholars regarding the reporting of the university services in support of the institutional activities. The fourth section highlights—in detail—the methodology adopted and the sample selected for the empirical analysis. The fifth section reports the results achieved and observations. The final section provides conclusions.

16.2 Performance Reporting in Public Organizations: Past Literature and Italian Legislative Initiatives

Since the advent of the new public management reform process (Hood 1991, 1995; Lapsley 2009; Pollitt and Bouckaert 2000), performance measurement has sparked renewed interest in the public sector, both within the academic and managerial communities. Some authors have mentioned its benefits, while others have highlighted the challenges that public organizations must face when trying to develop and use performance measurement procedures effectively (Behn 2003; De Bruijn 2002; Hatry 2002; Noordengraaf and Abma 2003; Van Thiel and Leeuw 2002).

Aside from the new public management, other paradigms, such as the “public value theory” and “public governance”, have also dealt with performance measurement from different perspectives. For the new public management, the interest of performance measurement mainly concerns the efficiency of public organizations from an intra-organizational viewpoint (Bouckaert and Halligan 2008). The “public value theory”, which was brought up by Moore (1995) before being developed over the past two decades (Williams and Shearer 2011), aims to rethink about how to create and manage (through an appropriate performance measurement system) efficiency, accountability and equity at the same time in the perspective of a co-creation value (Stoker 2006). Thus, the “public value theory” adopts a holistic approach and offers a broader way of measuring public performance. In detail, performance is multidimensional and covers service outputs, user satisfaction and outcomes, as well as trust and legitimacy (Kelly et al. 2002). The “public governance” emphasizes the active role of public organizations within other networks with regards to their level of interaction and accountability toward external stakeholders, while simultaneously paying attention to the public network performance and to communicating the results (Kickert 1997; Osborne 2010; Rhodes 1996). Consequently, the “public governance” also embraces a “plural and pluralist” (Osborne 2010) perspective towards a public performance that is focused on interorganizational relationships (i.e. networks), outcomes and stakeholder engagements (Schachter 1995).

Therefore, the definition of “what” a performance measurement system has to measure depends on the reason behind the need for such measurement; i.e., what needs and audiences must the information be able to satisfy. Once the measurement has been identified, a performance measurement system based on the practices, procedures, criteria and standards that govern the collection of data, its analyses and results, both qualitative and quantitative, must be implemented (Halachmi and Bouckaert 1996). Performance measurement is thus essentially the set of activities developed to gather information on performance (Van Dooren et al. 2010).

Of course, the usefulness of a performance measurement system mainly depends on the how the information provided will be used. The literature summarizes how the information on performance can potentially be used by dividing it into the following categories (Behn 2003; Bouckaert and Halligan 2008; Hammerschmid et al. 2013; Jansen 2008): *learning*, to improve management at strategic and operational levels; *steering and controlling*, to support managers in making decisions about the activities of their organizational units; *external accountability*, to enhance communication with the stakeholders as far as the results achieved are concerned. All of these uses highlight the fact that there is a strong connection between the performance measurement and the decision-making process and that both are needed to achieve the so-called “performance management”, i.e. “a type of management that incorporates and uses performance information for decision-making” (Van Dooren et al. 2010).

Performance management can be considered an evolution of performance measurement: indeed, “performance management precedes and follows performance measurement, in a virtuous spiral” (Lebas 1995). Despite some criticalities

concerning its implementation with complex contexts, such as public organizations, and which have already been underrated by scholars (Noordegraaf and Abma 2003; Calciolari 2009), performance management—through the use of performance measurement information—can “affect positive change in organisational culture, systems and processes” (Folan and Browne 2005) when the information is used for learning, steering and controlling and providing external accountability.

Thus, for example as far as the *learning* process is concerned, performance measurement provides useful information on how to improve the strategies and plans. In *steering and controlling* management, performance information improves the decision-making process and makes the organizational units and individual employees more accountable for the results achieved (Cavalluzzo and Ittner 2004). Finally, with regard to *external accountability* (the focus of this chapter), an effective performance measurement system will make the activities implemented more transparent and auditable, thus more complete in terms of fulfilling the level of responsibility toward the stakeholders, who can then in turn take action accordingly by influencing the public organizations’ decisions (Cavalluzzo and Ittner 2004; Talbot 2005; Yang and Holzer 2006). In this regard, it should be remembered that the effectiveness of *external accountability* is strongly connected to the definition of clear, univocal, non-contradictory, and quantifiable objectives in the planning phase. Indeed, as highlighted by several authors (Anselmi 1995; Borgonovi 2002), planning, control and external reporting are strictly linked and interdependent. Ambiguous objectives—often one of the distinguishing features of public organizations—are likely to generate a lack of clarity on “what” needs to be measured, thereby compromising the effectiveness of external reporting (Anselmi 2009).

In view of the performance measurement requirements mentioned above, in Italy, the national parliament introduced a mandatory performance measurement system for all public organizations in 2009, and that includes an evaluation of the managers and staff (Legislative Decree no. 150/2009; subsequent implementation of the resolutions and guidelines of the Evaluation, Transparency and Integrity Commission, also abbreviated as CiVIT and now referred to as the National Anti-Corruption Authority, ANAC). Under this decree, all public organizations are required to implement the so-called “Performance Management Cycle”, which includes the definition and assignment of objectives to be achieved, the identification of the necessary resources, the monitoring of ongoing initiatives, the implementation of any corrective actions, the use of reward systems based on merit evaluation criteria, and, finally, the reporting of results. This legislative initiative relies on the principles and tools that are already used by the public organizations, but also obliges the latter to publish new documents on their websites, such as a “Performance Measurement and Evaluation System” file to formalize the Performance Management Cycle’s model, phases, and timing, a “Performance Plan”, a multiyear programming tool and a “Performance Report” containing the results achieved.

The Performance Measurement and Evaluation System identifies “what” to measure in several areas of a public organization as a whole (organizational performance), but also in terms of the managers and staff (individual performance). As

far as organizational performance (the focus of this chapter) is concerned, the dimensions of the performance measurement defined by CiVIT (Resolutions no. 89/2010 and 104/2010) are: *satisfaction of the community's needs, implementation of plans and programs, user satisfaction with the activities and services, modernization and qualitative improvement, development of stakeholder relationships, efficiency in the use of resources, quality and quantity of the services provided and equal opportunity*. These dimensions of performance measurement are in line with the ones that are usually suggested in literature: *output* (“implementation of plans and programs”), *quantitative effectiveness* (“quantity of the services provided”), *qualitative effectiveness* (“user satisfaction with the activities and services”, “modernization and qualitative improvement”, “development of stakeholder relationships”, “quality of the services provided” and “equal opportunity”), *outcome* (“satisfaction of the community's needs”) and *efficiency* (“efficiency in the use of resources”).

To fully implement the new provisions, CiVIT explicitly suggests some models taken from the private sector: the performance tree, the balanced scorecard (Kaplan and Norton 1992), the performance prism (Neely et al. 2001) and the common assessment framework (Bovaird and Loffler 2003).

According to CiVIT, the minimum requirements that the performance measurement system needs to adopt are identified with reference to the characteristics formerly highlighted in Italian literature: a clear *definition of the performance objectives* to be achieved; the use of *outcome indicators*; the *specification of the connection* between the performance objectives, performance indicators (including outcome indicators) and targets; the *drafting of technical protocols* for each performance indicator to facilitate their implementation and to make them transparent (Anessi Pessina et al. 2016; Bergamin Barbato 1991; Brunetti 1979; Brusa and Zamproga 1991).

The formalization of these objectives involves the drafting of a Performance Plan, which is a three-year planning document that outlines—consistently with the resources available—the objectives that must be achieved, the performance indicators and the targets of the measurement of performance. This means that the performance measures do not stem from ministerial requirements, but rather from the objectives and indicators each public organization defined autonomously, as long as such measures are conducted according to the criteria established by the Performance Measurement and Evaluation System. CiVIT Resolution no. 112/2010, “How to structure and prepare the Performance Plan”, defines the contents of the Performance Plan and provides a logical map that defines the links between the institutional mandate, the mission, the strategic areas and objectives, the actions to be undertaken and the available or acquirable resources. The mandatory online publication of the Performance Plan keeps stakeholders informed and supports the accountability and transparency of performance.

As per the reporting of performance, the public organization must prepare the Performance Report to show and compare the results obtained during the previous year with the preset objectives. The report should be approved by June of the following year. It is drafted according to the resources available; any gaps between

the objectives and results are acknowledged and their causes and corrective measures must be described. This report is addressed to the policymakers, public managers, external auditors, citizens and all parties involved. According to the legislation, the Performance Report should meet the information needs of the multiple internal and external stakeholders of the organization and should highlight the critical variables that can be influenced by the decisions of internal stakeholders, but also the benefits of the external stakeholders in order to implement a comprehensive information framework aimed to control the organization's activities.

In other words, the Performance Report must be able to respond to the very different information needs of the different parties involved, while remaining a concise and easy to understand document aimed at a broad and varied audience (though containing annexes with more detailed information). It should be remembered that the structure and form of the document are important aspects of its effectiveness as a communication tool.

Legislative Decree no. 150/2009 does not specify a compulsory structure for this document, but CiVIT issued Resolution no. 5/2012 to define its contents: "introduction and table of contents", "abstract of relevant information for citizens and other external stakeholders" (including basic information on the organization's external and internal contexts, its outputs and outcomes, and critical issues and opportunities), "results achieved and variations", "resources, efficiency, and economy", "results achieved in terms of gender policy", and "description of the preparation process of the report" (phases, actors, timing, and responsibilities, as well as strengths and weaknesses of the Performance Management Cycle). The Performance Report, as regulated by CiVIT, is thus a document that must meet certain formal requirements to make sure it can be used to evaluate the individual and organizational performance. Furthermore, the document must satisfy the information needed by the steering and control management functions.

With the support of CiVIT, the legislator thus recalls the managerial principles needed for the development of performance management. In fact, as required by the literature, he will highlight the need to provide all of the information related to the performance (results achieved and resources used, a comparison with what was planned, an analysis of the deviations, a temporal and spatial comparison) in order to better understand the management's progress compared to the preset objectives and to make evaluations that can be used in future planning (Hatry 2006; Van Dooren et al. 2010).

The set of objectives (evaluative and informative) and recipients (internal and external) that characterize a Performance Report makes this document extremely important, but its complex structure and the large amount of information it needs to cover may prevent it from being an adequate framework as far as the communication of the activities and the performance of public organizations are concerned.

Like all public organizations, public universities are also required (by the CiVIT Resolution 9/2010) to implement a performance measurement system, albeit autonomously and with their own organizational methods (Dal Molin et al. 2016; Zuccardi Merli and Bonollo 2013). However, it should be emphasized that this provision only applies to nonteaching staff and therefore only in support of the

research, teaching, and third mission activities (although this changed in 2016, with new ministerial guidelines; public universities now have to include results for all of their activities). Supporting services tend to be neglected in literature on performance measurement, as it traditionally focuses on measuring research and teaching activities.

In order to provide a theoretical framework for other empirical investigations, we will highlight the main scholarly contributions on the performance measurement of the activities in support of research, teaching and the third mission of universities in the following section.

16.3 University Performance Measurement for the Services in Support of Institutional Activities

Performance measurement is not a new concept for Italian public universities, as they have had to refer—for years already—to the Ministry of Education, Universities and Research with their results in research, teaching and the third mission, and sometimes even for the allocation of financial resources.

Alongside this requirement of performance measurement in response to institutions, different forms of accountability have been developed over time and created a series of very complex performance measurement systems. Literature has long identified the peculiarities of universities, compared to other kinds of organizations, and revealed the particular challenges of performance measurement for that specific setting—the intangibility of the teaching and research activities, the typical matrixes of the organizational structure, difficulties in the attribution of outcomes, the high number of stakeholders involved, etc. (Broadbent 2007; Cosenz and Bianchi 2013; Rabovsky 2014; Vakkuri and Mecklin 2003).

Studies on university performance measurement often focus on measuring the performance of research and teaching of the contributions of specific countries (Beasley 1995; Riccaboni 2003; Taylor 2001) by making international comparisons (Lewis 2014; Pettersen 2015) or by evaluating related issues (Borgonovi and Giordano 2007; Rebora 2003; ter Bogt and Scapens 2012; Turri 2005). However, the performance measurement contributions related to supporting services are still limited, and often involve surveys of the overall administrative quality of the services or specific projects in the Italian context (Arena et al. 2009; Arnaboldi et al. 2007; Catalano 2004; Cugini and Pilonato 2006; Rebora 2003). The same is also true in other contexts, such as the United Kingdom (Oldfield and Baron 2000; Smith et al. 2007) and Australia (Waugh 2002). Despite the almost complete lack of performance measurement studies on supporting services, such are undoubtedly important; their activities can greatly influence the overall quality of university work, especially in terms of efficiency (as they can wear out the resources if not managed in a rational way) and stakeholder perception.

Supporting services include very different types of activities and involve stakeholders with various needs and characteristics, such as teachers, students, companies and public organizations, among others. Precisely because of this heterogeneity, which makes the measurement more complex, some authors begin by classifying these services into homogeneous categories.

Rebora (2003) focuses on the types of activities and thus classifies them into *technical and logistical* support services (catering, cleaning, maintenance, surveillance); *coordination and administration* support services, which are provided to institutional bodies (procurement and contracts, fundraising, research, project management, analysis, and studies), or others in direct support of *teaching and research* (secretariats, libraries, information services, workshops, tutoring students, internship and international exchange organizations, postgraduate placements). Rebora suggests that each of these services and specific activities have different key factors in terms of value creation and which can either be valuable for the various organizations of other fields (e.g. for private companies, such as the *technical and logistical* support services) or specifically aimed at universities (e.g. relationships between the institutional bodies and their professors and students, for the other two types).

Cugini and Pilonato (2006) focus on the users of support services and break down the support services into two main categories, accordingly: *internally oriented services*, provided to organizational units within the university (general administrative services, human resources, IT systems), and *externally oriented services* in direct support of research and teaching (libraries, laboratories, student secretariats).

With regard to this user-based breakdown, Arnaboldi et al. (2007), as part of the “Good Practice Project” that involved a total of 34 Italian public universities from 1998 to 2007, present a process-based scheme according to the input received from the participants of the project and which divides the support services into *student support services*, *research support services*, *human resources*, *accounting*, *information services*, *procurement*, *logistics* and *library services*.

The aforementioned authors then respectively assess the activities (Rebora 2003), the critical issues connected to the identification of appropriate measures (Cugini and Pilonato 2006) and the cost of the services (Arnaboldi et al. 2007). To this date however, an aspect has been neglected: the methods that are actually adopted by universities to measure the supporting activities and which, as already mentioned, Italian universities must introduce as result of the provisions of Law 15/2009.

16.4 The Methodology Applied

Our investigation considers the entire roster of Italian public universities and is based on their Performance Reports over 4 years (2011–2014); the same has been broken down in two phases.

In the first phase, we checked which of the 66 Italian public universities had published Performance Reports on their websites, or rather which had completed the last phase of the Performance Management Cycle to fulfil the transparency requirements set by law.

The second phase focused on the contents of the 209 Performance Reports identified in the previous phase. First, we checked the Performance Reports for compliance with the CiVIT requirements (Resolution 5/2012). Then, we proceeded with a more detailed analysis of their content to verify the presence of performance information (which scholars identified as being important and in line with the managerial principles), the frequency of organizational performance indicators and the relevant support service by type.

The first set of information we analysed in the second phase is reported in Table 16.1; the presence/absence of such information was evaluated with dichotomous answers (yes/no) while focusing on the following areas: results achieved (year-end value, percentage of achievement, or declaration of failure/partial achievement/total achievement); comparing the activities carried out to the results obtained and to what was reported in the Performance Plan; comparing the results obtained to the human, technical, and financial resources used; comparing the allocation of responsibilities to the results obtained; identifying the causes of the variations compared to the preset objectives and identifying the corrective initiatives; comparing over time and space with other universities.

The subsequent and more thorough analysis based on the type of support service and performance indicator did not consider all 209 performance reports for 2011–2014. In fact, it only focused on the 2014 reports of the 39 universities that had

Table 16.1 Grid to verify the implementation of the managerial principles

Performance report
Contains information about the results achieved (year-end value, percentage of achievement, declaration of failure/partial achievement/total achievement)
Describes the activities carried out and the results obtained based on generic connections to the Performance Plan
Makes explicit and detailed comparisons between the activities carried out, the results obtained, and the objectives of the Performance Plan
Establishes links between the results obtained and:
The organizational units in charge
The human resources used
The technical resources used
The financial resources used
Explains the causes of the variation (+ or –) compared to the objectives set
Describes the corrective initiatives adopted
Shows the time trend of the results obtained
Provides information relating to benchmarking

ensured the continuity of the Performance Management Cycle over time by drafting and publishing an annual Performance Report from 2011 to 2014. The decision to consider only the 2014 reports was made to ensure a synchronic comparison; no trend analysis was conducted over time, under the assumption that the most recent university reports were drafted in an efficient and reliable manner considering these institutions' consolidated experience.

The analysis was based on the grid shown in Table 16.2, which also reports the type of support services and performance indicators we adopted. This classification considered the proposals of scholars and the primary guidelines of CiVIT, as described in the previous sections.

The performance indicators are as follows: *output, quantitative and qualitative effectiveness, outcomes, efficiency, expenses/costs, and revenues*. The support services are divided into *internally oriented support services* (technical and logistical support services, administration and coordination services) and *externally oriented support services* (services in support of research and teaching, patents and spin-offs, sports and student associations). In some cases, a more detailed level was prepared to highlight specific activities in support of the institutional mission.

16.5 Main Results and Discussion

The institutional website survey revealed that 62 out of 66 Italian public universities (93.94%) had published Performance Reports on their websites at least once during the period under investigation. In particular, 43 Performance Reports were published in 2011, 50 in 2012, 59 in 2013, and 57 in 2014. Only four universities never published a Performance Report.

The documentary analysis initially considered 209 Reports. Almost all of the documents complied with the Performance Report structure defined by CiVIT in Resolution no. 5/2012. Only six universities provided non-conforming documents; for example, the University of Venice Ca' Foscari prepared a "University Report" that contained a Performance Report, a Sustainability Report and the rector's mandatory report on research, teaching, and technology transfer (Law 1/2009).

As for the documents published by the remaining universities, they were all improperly drafted as there apparently was a lack of attention to the sections concerning the "resources, efficiency, and economy" and the "strengths and weaknesses of the performance management cycle" (which are both mandatory, according to the CiVIT Resolution).

In the case of resources, efficiency, and economy, the link between the financial statements and performance indicator results is often absent or restricted to financial statement extracts, notes on the transition to the new accrual accounting system, traditional expenditure indicators and indicators for the personnel expenses and indebtedness as required by the Ministry of Education, Universities and Research. The lack of any link between the non-financial performance indicators and the financial resources is sometimes justified in the reports by the fact that the

Table 16.2 Classification of the services in support of research, teaching and the third mission and their relevant performance indicators

Performance indicators	Output		Quantitative effectiveness		Qualitative effectiveness		Outcome		Efficiency		Expenses/costs and revenues	
	Nonquantitative	Quantitative	Nonquantitative	Quantitative	Nonquantitative	Quantitative	Nonquantitative	Quantitative	Nonquantitative	Quantitative	Nonquantitative	Quantitative
Support services												
Internally oriented												
Technical and logistical support services												
Building, residential services, teaching and research facilities, safety at work												
Catering												
Cleaning												
Surveillance												
Administration and coordination services												
Human resources												
Planning and controlling, accounting, administrative simplification, management of subsidiaries												
Procurement and contracts												
IT systems												
Legal activities, support to institutional bodies, communication, websites												
Externally oriented												
Research and teaching support services												
Research project management, fundraising, internationalization												
Libraries, laboratories, botanical gardens, museums												
Student secretariats (including activities supporting post graduation)												
Tutorship, placement, organization of internships and international exchanges, scholarships and other benefits												
Patents and spin-off support services												
Sports and student association support services												

institution did not adopt the accrual accounting system since traditional public accounting would hinder the relationship between the goals and resources.

As for the strengths and weaknesses of the performance management cycle, of the 209 documents examined, a significant number (79) did not indicate the institution's main strengths and weaknesses, which account for one of the items of performance measurement process. In fact, this was the case for 53% of the documents in 2011, 44% in 2012, 34% in 2013, and 25% in 2014. This might indicate a bureaucratic behaviour and reluctance to use the opportunity provided by the requirement to implement a performance measurement system aimed to improve and support the decision-making processes.

A subsequent analysis was carried out and reported on the grids drafted by the authors (Tables 16.1 and 16.2). The objective was to verify whether the content of the documents could be considered to have been inspired by managerial principles, to determine which support services had actually been subjected to reporting requirements and to identify which types of performance indicators were measured.

The results for the 209 Performance Reports show that, in terms of the areas of activities subjected to reporting, most of the universities (54 of 66) had reported on activities related to research, teaching, and the third mission. Although the normative provisions are only concerned with support activities (until the changes in 2016), universities tend to project performance measurement on to their overall activities to provide an overview of their work rather than to adopt minimal reporting rules that only cover the activities of support.

With regard to the managerial principles used in these documents, only three universities, each across all of the years under investigation, did not present the planned objectives (already published in the Performance Plan) that were intended to be compared to the results achieved; in other words, their Performance Reports only indicate the final results achieved or the degree to which the goal was achieved. The remaining institutions mainly published Performance Reports containing a description, at times very detailed, of the activities carried out and of the results achieved, while only providing a generic link to the content of the Performance Plan; of all the reports published during this period, only about 25% of universities carried out explicit and detailed comparisons between the objectives mentioned in the Performance Plan and the results achieved.

The verification of the requirements in Table 16.1 highlighted many other shortcomings, such as the lack of any variation analysis (it was present only in 20 Performance Reports, which were published by 13 universities) and of having identified the measures needed to address failure in term of reaching the objectives (corrective actions were laid out in only one Performance Report out of 209).

Overall, the critical issues and opportunities that arose during the timeframe were identified, but almost all of the documents examined refer to the activities of the university as a whole and do not address support services separately nor, with the exception of one case, the achievement or failure of specific objectives. Also, there was a general lack of information that makes it impossible to make comparisons over time, to identify standards of reference compared to other universities and to establish links between the results obtained and the human, technical, and

financial resources used to obtain them (that is, a lack of explicit coordination with the financial statements). Only two universities compare their indicators to the national average indicators proposed by the “Good Practice Project”, which evaluates the perceived quality of the supporting services. Thus, the Performance Report seems to be seen as a mere *ex post* reporting document that neither tries to suggest corrective actions nor to impact the subsequent university planning.

As for types of support services, as mentioned, only 39 Performance Reports for 2014 were considered. They reveal a variety of activities that universities consider appropriate in terms of measurement and monitoring. In this regard, the obligation to identify and measure individual performance, which is connected to an incentive/bonus systems for the managers and staff, certainly has been a strong stimulus in favour of the performance measurement of the universities’ activities. It should, however, be noted that, in several cases, Performance Reports consider organizational performance as marginal and that they thus end up attributing greater importance (based on the amount of information reported) to individual performance.

Overall, we assigned different degrees of importance to different types of supporting activities based on the number of organizational performance indicators identified (Table 16.3).

There generally is a higher number of performance indicators for services that are addressed to students (21% of the total on the average) and which can be broken down into “tutorship, placement, organization of internships, international exchanges, scholarships, and other benefits” (15% on the average), as well as “student secretariats” (6% on the average). This seems to indicate that universities regard their students as the most important stakeholders, even if other services—such as “catering” (i.e., the cafeteria) and “cleaning”—could undoubtedly affect quality from a student’s point of view. However, such are not taken into account (0% on the average).

There is also a relatively high number of performance indicators relating to “legal activities, support to institutional bodies, communication, and websites” (15% on the average), as well as to “IT systems” (13% on the average). In the first case, the high number is a result of the legal provisions on transparency and anti-corruption, whereas in the second case, the numbers reflect the increasing efforts to provide administrative procedures and online services.

Attention paid to the “planning and controlling, accounting, administrative simplification, and management of subsidiaries” (13% on the average), as well as to “human resources” (12% on the average) is closely connected to the fact that times are changing in terms of management and accounting information systems (Zuccardi Merli and Bonollo 2013), thus resulting in the need to develop training activities for the managers and staff.

Based on our findings, the apparently less important supporting services included “research project management, fundraising, and internationalization” (7% on the average), “libraries, laboratories, botanical gardens, and museums” (5% on the average), “activities in support of patents and spin-offs” (1% on the average) and “activities in support of sports and student associations” (0% on the average).

Moving on to the type of performance indicators, the results reported in Table 16.4 show that most (79% on the average) are related to the type and

Table 16.3 Performance indicators classified by type of support services

Universities	Technical and logistical support service			Administration and coordination services				Research and teaching support services				Tutorship, placement, internship and international exchanges, scholarship and other benefits	Patents and spin-off support services	Sports and student association support services
	Catering	Surveillance	Cleaning	Personnel management	Planning and controlling, accounting, administrative simplification, management of subsidiaries	Procurement	IT systems	Legal activities, support to institutional bodies, communication, websites	Research project management, fundraising, internationalization	Libraries, laboratories, botanical gardens, museums	Student secretariats (including activities supporting post graduation)			
	% out of total	% out of total	% out of total	% out of total	% out of total	% out of total	% out of total	% out of total	% out of total	% out of total	% out of total	% out of total	% out of total	% out of total
1	9	0	0	15	24	0	21	8	6	1	6	11	0	0
2	9	0	0	6	12	6	12	10	11	2	13	19	0	0
3	14	0	0	11	4	0	11	4	11	4	0	36	7	0
4	7	0	0	15	8	0	14	18	6	8	10	13	0	0
5	9	0	2	9	8	0	9	28	0	17	0	16	2	0
6	7	0	1	9	15	1	9	16	7	7	3	22	0	0
7	0	0	0	0	27	0	0	0	0	17	0	53	3	0
8	33	0	0	8	14	0	14	10	4	0	0	12	6	0
9	7	0	0	9	12	0	17	14	5	7	9	18	1	0
10	6	0	0	17	19	3	11	31	0	11	0	3	0	0
11	0	0	0	0	3	0	9	3	31	6	6	41	0	0
12	6	0	0	8	29	2	8	40	0	2	4	0	0	0
13	6	0	0	6	12	9	15	26	4	4	7	6	1	3
14	5	0	0	16	16	0	21	11	5	11	0	16	0	0
15	17	0	0	30	4	4	17	0	9	0	0	17	0	0
16	23	0	0	16	19	2	9	23	0	0	5	2	0	0
17	6	1	0	1	9	0	6	7	7	6	18	43	1	0
18	7	0	0	14	7	0	0	6	14	0	21	29	0	0
19	16	0	0	7	11	0	21	13	11	2	7	11	0	0
20	15	0	0	22	14	0	18	9	2	3	11	6	0	0
21	8	0	0	0	12	4	4	20	8	0	0	40	0	4
22	12	0	2	11	15	2	12	17	2	3	9	12	3	0
23	0	0	0	20	20	0	20	20	20	0	0	0	0	0

(continued)

Table 16.3 (continued)

Universities	Technical and logistical support service			Administration and coordination services				Research and teaching support services				Patents and spin-off support services	Sports and student association support services
	Catering	Surveillance	Cleaning	Personnel management	Planning and controlling, accounting, administrative simplification, management of subsidiaries	Procurement	IT systems	Legal activities, support to institutional bodies, communication, websites	Research project management, fundraising, internationalization	Libraries, laboratories, botanical gardens, museums	Student secretariats (including activities supporting post graduation)		
	% out of total	% out of total	% out of total	% out of total	% out of total	% out of total	% out of total	% out of total	% out of total	% out of total	% out of total	% out of total	% out of total
24	7	0	0	16	20	4	9	4	13	11	14	2	0
25	11	0	0	27	15	2	11	19	2	2	1	11	0
26	0	0	0	3	3	0	24	13	8	42	0	8	0
27	12	0	0	12	12	6	19	17	4	2	13	4	0
28	16	0	0	13	7	8	17	12	4	0	12	5	7
29	5	0	0	10	0	0	0	20	15	5	5	35	0
30	4	0	0	19	11	0	37	0	11	0	0	11	7
31	10	0	0	20	9	2	22	10	10	4	3	8	1
32	33	0	0	0	0	8	0	42	0	0	17	0	0
33	50	0	0	50	0	0	0	0	0	0	0	0	0
34	4	0	0	11	7	7	9	35	10	5	7	3	1
35	21	0	0	4	25	4	11	4	4	18	7	4	0
36	0	0	0	0	17	0	17	0	17	0	0	50	0
37	0	0	0	7	11	7	22	22	11	0	15	4	0
38	9	0	0	6	29	3	9	24	6	0	0	15	0
39	2	0	0	2	12	10	21	26	14	7	0	5	0
Mean	10	0	0	12	13	2	13	15	7	5	6	15	1
Median	7	0	0	10	12	1	12	13	6	3	5	11	0
Minimum				0	0	0	0	0	0	0	0	0	0
Maximum	50	1	2	50	29	10	37	42	31	42	21	53	7
Standard dev.	10	0	0	10	8	3	8	11	7	8	6	15	2

Table 16.4 Performance indicators classified by types

Universities	Output			Quantitative effectiveness			Qualitative effectiveness			Outcome			Efficiency			Expenses/cost revenues		
	Non-quantitative % out of total	Quantitative % out of total	% out of total	Non-quantitative % out of total	Quantitative % out of total	% out of total	Non-quantitative % out of total	Quantitative % out of total	% out of total	Non-quantitative % out of total	Quantitative % out of total	% out of total	Non-quantitative % out of total	Quantitative % out of total	% out of total	Non-quantitative % out of total	Quantitative % out of total	% out of total
1	52	28	80	0	11	11	0	0	9	9	0	0	0	0	0	0	0	0
2	12	17	29	0	3	3	54	12	65	0	1	1	0	0	0	0	1	1
3	4	22	26	0	0	0	0	65	65	0	9	9	0	0	0	0	0	0
4	0	28	28	0	13	13	0	34	34	0	10	10	0	1	1	0	14	14
5	50	27	76	0	5	5	6	11	17	0	1	1	0	1	1	0	0	0
6	87	10	97	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0
7	0	86	86	0	0	0	0	5	5	0	9	9	0	0	0	0	0	0
8	34	30	64	0	0	0	0	22	22	0	2	2	0	0	0	0	12	12
9	73	22	95	0	3	3	0	3	3	0	0	0	0	0	0	0	0	0
10	61	14	75	0	11	11	0	11	11	0	0	0	0	0	0	0	3	3
11	72	25	97	0	0	0	0	0	0	0	0	0	0	0	0	0	3	3
12	71	29	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	1	96	97	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0
14	47	5	53	0	0	0	0	37	37	0	11	11	0	0	0	0	0	0
15	87	0	87	0	0	0	13	0	13	0	0	0	0	0	0	0	0	0
16	9	86	95	0	5	5	0	0	0	0	0	0	0	0	0	0	0	0
17	6	6	12	0	1	1	1	52	54	0	0	0	0	0	0	0	33	33
1	64	29	93	0	0	0	7	0	7	0	0	0	0	0	0	0	0	0
19	50	46	96	0	0	0	0	4	4	0	0	0	0	0	0	0	0	0
20	100	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	8	84	92	0	0	0	0	0	0	0	4	4	0	0	0	0	4	4
22	71	20	91	0	0	0	0	9	9	0	0	0	0	0	0	0	0	0
2	20	40	60	0	0	0	0	40	40	0	0	0	0	0	0	0	0	0
24	0	61	61	0	0	0	0	39	39	0	0	0	0	0	0	0	0	0
25	37	56	93	0	0	0	0	3	3	0	2	2	0	0	0	0	1	1
26	16	45	61	0	0	0	5	29	34	0	5	5	0	0	0	0	0	0
27	56	42	98	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2
28	80	19	99	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1

(continued)

Table 16.4 (continued)

Universities	Output			Quantitative effectiveness				Qualitative effectiveness				Outcome				Efficiency				Expenses/cost revenues			
	Non-quantitative	Quantitative		Non-quantitative	Quantitative		Non-quantitative	Quantitative		Non-quantitative	Quantitative		Non-quantitative	Quantitative		Non-quantitative	Quantitative		Non-quantitative	Quantitative			
		% out of total	% out of total		% out of total	% out of total		% out of total	% out of total		% out of total	% out of total		% out of total	% out of total		% out of total	% out of total		% out of total	% out of total	% out of total	% out of total
29	30	65	95	0	5	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
30	30	56	85	0	0	0	0	0	7	7	0	0	0	0	0	0	0	0	0	0	0		
31	37	61	98	0	0	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0		
32	17	67	83	0	0	0	0	0	8	8	0	0	0	0	0	0	0	0	0	0	8		
33	0	50	80	0	0	0	0	0	50	50	0	0	0	0	0	0	0	0	0	0	0		
34	45	39	84	0	0	0	2	12	14	14	0	1	0	0	0	0	0	0	0	0	1		
35	100	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
36	0	50	50	0	0	0	0	17	17	17	0	33	0	0	0	0	0	0	0	0	0		
37	59	26	85	0	0	0	11	11	11	11	0	4	0	0	0	0	0	0	0	0	0		
38	71	26	97	0	0	0	0	3	3	3	0	0	0	0	0	0	0	0	0	0	0		
39	93	5	98	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2		
Mean	42	36	79	0	1	1	3	12	15	15	0	3	0	0	0	0	0	0	0	0	2		
Median	45	29	87	0	0	0	0	4	7	7	0	0	0	0	0	0	0	0	0	0	0		
Minimum	0	0	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
Maximum	100	96	100	0	13	54	65	65	65	65	1	33	0	33	0	1	0	0	0	0	33		
Standard deviation	32	26	24	0	3	9	17	17	19	19	0	6	0	6	0	0	0	0	0	0	6		

importance of the activities, as well as to the importance of the results achieved. Typically, these are nonquantitative indicators of dichotomous nature (either “done” or “not done”), although some universities report the degrees of achievement, compared to the preset targets, in percentages (without specifying how such percentages are determined however) or define concrete quantitative targets. On the average, 15% of the indicators are made up of qualitative effectiveness indicators, usually based on the quality that is perceived, as assessed through questionnaires (using the Likert scale) or by monitoring the timing of services provided.

There are fewer indicators for the outcomes (3% on the average), quantitative effectiveness (1% on the average), and financial aspects (expenses/costs and revenues) (2% on the average); those related to efficiency are almost inexistent.

Such evidence, which shows a clear preference toward the output indicators, probably reflects the stakeholders’ greater propensity to interpret what data and supporting activities are easiest to deal with compared to the difficulty/burden of finding the information needed to create other types of indicators (consider, for example, the complexity involved in defining a cause–effect relationship between the support services and the improvements of institutional research, teaching, and third mission activities).

Based on the analysis, whose results were analysed in this section, it appears that the innovative measures launched by the parliament and CiVIT have improved the reporting operations of public universities as far as their supporting activities are concerned, something which had traditionally been neglected in the past. However, there is still much room for improvement, especially in terms of analysing the results obtained, of developing outcome measures and of coordinating the support service indicators with the financial statements to further assist the decision-making process and develop performance management.

16.6 Conclusions

In public organizations, performance reporting developed over time and assumed different connotations with the growing number of theories of the new public management, “public value” and “public governance”.

The performance measurement system of a public organization does indeed depend on the information requirements that need to be satisfied. Such needs generate in turn the so-called needs of performance management, i.e. the development of a management model that incorporates and uses the information on performance to guide the decision-making process according to the managerial principles.

In light of the need to measure performance, in 2009, the Italian legislator introduced and regulated a system designed to measure and evaluate performance in order to help all public organizations operate on the basis of planned objectives, whose results can be monitored and reported. More specifically, the legislator requires the inclusion of the planned objectives within the *Performance Plan* and a

need to report the amount of progress within the Performance Report to meet the information requirements of a variety of stakeholders, both within and outside of the public organization, for the future planning needs.

In particular, with regard to the organizational performance of the public organization as a whole and which is analysed in this chapter with specific reference to the activities in support of research, teaching and third mission of the public universities, the measuring ranges are connected to the traditional dimensions of the performance provided by scholars (activity, qualitative and quantitative effectiveness, outcomes and efficiency) and required by CiVIT within the guidelines in support of the implementation of the Legislative Decree 150/2009.

In this context, the objective of the work was to verify the compliance of the Performance Report according to the Legislator's provisions and to the literature, thus by respecting the managerial principles in order to confirm which performances were actually measured and reported to the university.

The analysis highlights the large number of public universities that have applied the provisions on performance measurement set out in Law 15/2009—more than 90% have published at least one Performance Report from 2011 to 2014. So, legal enforcement has successfully driven universities to measure the performances, at least in the initial phase of reform.

It should be remembered that these Performance Reports often do not include the information retained important by scholars in the next programming cycle—for instance, the lack of an adequate variation analysis—and there is insufficient information on the resources used. Therefore, the regulatory imposition did not ensure the implementation of the requirements needed for the development of performance management. In many cases, the reform does indeed seem to have only been implemented for compliance purposes rather than for managerial ones. In order to ensure that the information is suitable enough to support a decision-making process according to the managerial principles, more support is needed from the National Agency for the Evaluation of the University and Research System (which currently performs the duties that were once executed by CiVIT/ANAC) to help institutions meet these requirements and to adopt principles of transparent management. More commitment from the Ministry of Education, Universities and Research could also lead to incentives in this regard, and training for managers will help demonstrate the potential of Performance Reports and secure buy-ins at their level.

The documents examined showed evidence that the performance measurement of the services addressed to students (i.e., in support of the teaching activities) are of greater importance and interest than the performance measurement of the services in support of research and the third mission, which are largely neglected. Some attention is paid to internal services such as the IT systems, accounting information system and activities that are closely related to the institutional bodies and general manager. This is partly justified from the time of transition (organizational and accounting information system reforms) that characterized Italian public universities during the period under investigation.

Regarding the types of indicators, the suggestions provided by the scholars and CiVIT were only partly reproduced. Almost 80% of the performance indicators are output indicators, thus inadequate attention is paid to all others, especially in terms of outcome, qualitative effectiveness, and efficiency, whose indicators could help improve the management of the supporting services of the institutional activities in terms of impacts, perceived quality and the use of the resources. As already mentioned, the effective and efficient image of a university depends on the research, teaching, and third mission, but also on how it provides supporting services and on how they are perceived. It is thus clear that the performance concept is interpreted in a restrictive manner, without accepting, if not minimally, the recommendations of the new public management's scholars (there are practically no efficiency indicators) on the "public value theory" (few outcome indicators and no trustworthy and legitimate reference) and on "public governance" (no reference within the qualitative effectiveness indicators on the stakeholder relationships).

In conclusion, a few years after the legislation on performance measurement system went into effect, we noticed that the universities increasingly began to use the documents that were introduced by the legislator for this purpose. On the other hand, by developing the analysis of the Performance Report, we had to demonstrate how the contents of such documents still do not fully meet the requirements of the managerial principles. Hence, the legislation has been an insufficient, although important, instrument of managerial innovation. In the future, the findings of this study can further be expanded by investigating the institutions' actual decision-making process with regards to the information analysed in this document. In this way, we could offer suggestions to promote the effective implementation of managerial innovations within public universities while making sure that they address the transparency and legal requirements of the managerial principles.

References

- Anessi Pessina, E., Barbera, C., Sicilia, M., & Steccolini, I. (2016). Public sector budgeting: A European review of accounting and public management journals. *Accounting, Auditing & Accountability Journal*, 29(3), 491–519.
- Anselmi, L. (1995). *Il processo di trasformazione della pubblica amministrazione*. Torino: Giappichelli.
- Anselmi, L. (2009). Amministrazioni da valutare, strumenti per misurare, attività da rendicontare: la multidimensionalità della performance. In: CNEL (Ed.), *La misurazione delle performance nelle pubbliche amministrazioni*. Roma: CNEL.
- Arena, M., Arnaboldi, M., Azzone, G., & Carlucci, P. (2009). Developing a performance measurement system for university central administrative services. *Higher Education Quarterly*, 63(3), 237–263.
- Arnaboldi, M., Azzone, G., & Carlucci, P. (2007). Misurare le performance nel settore pubblico: un'applicazione alle università. In A. Cugini (Ed.), *La misurazione della performance negli Atenei*. Milano: Franco Angeli.
- Beasley, J. E. (1995). Determining teaching and research efficiencies. *Journal of the Operational Research Society*, 46(4), 441–452.

- Behn, R. D. (2003). Why measure performance? Different purposes require different measures. *Public Administration Review*, 63(5), 586–606.
- Bergamin Barbato, M. (1991). *Programmazione e controllo in un'ottica strategica*. Torino: Utet.
- Bianchi, C. (2016). *Dynamic performance management*. Heidelberg: Springer.
- Borgonovi, E. (2002). *Principi e sistemi aziendali per le pubbliche amministrazioni*. Milano: Egea.
- Borgonovi, E., & Giordano, F. (2007). La valutazione dell'Università: aspetti sistematici ed operativi. In A. Cugini (Ed.), *La misurazione della performance negli atenei. Logiche, metodi, esperienze*. Milano: Franco Angeli.
- Bouckaert, G., & Halligan, J. (2008). *Managing performance: International comparisons*. New York: Routledge.
- Bovaird, T., & Löffler, E. (Eds.). (2003). *Public management and governance*. London: Routledge.
- Broadbent, J. (2007). If you can't measure it, how can you manage it? Management and governance in higher educational institutions. *Public Money & Management*, 27(3), 193–198.
- Brunetti, G. (1979). *Il controllo di gestione in condizioni ambientali perturbate*. Milano: Franco Angeli.
- Brusa, L., & Zamprognà, L. (1991). *Pianificazione e controllo di gestione*. Milano: Etas Libri.
- Calciolari, S. (2009). Gestione della pubblica amministrazione basata sui risultati: criticità, strumenti e prospettive. *Azienda pubblica*, 4, 645–664.
- Catalano G. (Ed.). (2004). *Valutare le attività amministrative delle Università: aspetti metodologici e buone pratiche*. Bologna: Il Mulino.
- Cavalluzzo, K. S., & Ittner, D. C. (2004). Implementing performance measurement innovations: evidence from government. *Accounting, Organization and Society*, 29(3/4), 243–267.
- Cosenz, F., & Bianchi, C. (2013). Designing performance management systems in academic institutions: A dynamic performance management view. *ASPA conference*, New Orleans.
- Cugini, A., & Pilonato, S. (2006). La misurazione della performance degli Atenei: uno schema per l'analisi dello stato dell'arte. *Azienda Pubblica*, 2, 197–215.
- Dal Molin, M., Turri, M., & Agasisti, T. (2016). New public management reforms in the Italian Universities: Managerial tools, accountability mechanisms or simply compliance? *International Journal of Public Administration*. Published online 31 Mar 2016.
- De Bruijn, H. (2002). *Managing performance in public sector*. London: Routledge.
- Folan, P., & Browne, J. (2005). A review of performance measurement: Toward performance management. *Computers in Industry*, 56, 663–680.
- Halachmi, A., & Bouckaert, G. (1996). Introduction: Performance appraisal and Rubik's cube. In A. Halachmi & G. Bouckaert (Eds.), *Organizational performance and measurement in the public sector*. Westport: Quorum Books.
- Hammerschmid, G., van de Walle, S., & Stimac, C. (2013). Internal and external use of performance information in public organizations: results from an international survey. *Public Money & Management*, 33(3), 261–268.
- Hatry, H. P. (2002). Performance measurement: Fashion and fallacies. *Public Performance & Management Review*, 25(4), 352–358.
- Hatry, P. H. (2006). *Performance measurement: Getting results*. Washington: The Urban Institute Press.
- Holzer, M., & Halachmi, A. (1996). Measurement as a means of accountability. *International Journal of Public Administration*, 19(11/12), 1921–1944.
- Hood, C. (1991). A public management for all seasons. *Public Administration*, 6(3), 3–19.
- Hood, C. (1995). The new public management in the 1980s: Variations on a theme. *Accounting, Organizations and Society*, 20(2/3), 93–109.
- Jansen, E. P. (2008). New public management: Perspectives on performance and the use of performance information. *Financial Accountability & Management*, 24(2), 169–195.
- Johnsen, A. (2005). What does 25 years of experience tell us about the state of performance measurement in public policy and management. *Public Money & Management*, 25(1), 9–17.
- Kaplan, R. S., & Norton, D. P. (1992). The balanced scorecard: Measures that drive performance. *Harvard Business Review*, 70(1), 71–79.

- Kelly, G., Mulgan, G., & Muers, S. (2002). *Creating public value: An analytical framework for public service reform*. London: Strategy Unit, Cabinet Office.
- Kickert, W. J. M. (1997). Public governance in the Netherlands: An alternative to Anglo-American 'Managerialism'. *Public Administration*, 75(4), 731–752.
- Lapsley, I. (2009). New public management: The cruellest invention of the human spirit? *Abacus*, 45(1), 1–21.
- Lebas, M. J. (1995). Performance measurement and performance management. *International Journal Production Economics*, 41, 23–35.
- Lewis, J. M. (2014). Research productivity and research system attitudes. *Public Money & Management*, 34(6), 417–424.
- Moore, M. (1995). *Creating public value: Strategic management in government*. Cambridge, MA: Harvard University Press.
- Mussari, R. (2001). Misurazione e valutazione della performance: un'analisi critica. In A. Lombroso (Ed.), *Il controllo di gestione negli enti locali*. Maggioli: Rimini.
- Neely, A., Adams, C., & Crowe, P. (2001). The performance prism in practice. *Measuring Business Excellence*, 5(2), 6–12.
- Noordegraaf, M., & Abma, T. (2003). Management by measurement? Public management practices amidst ambiguity. *Public Administration*, 81(4), 853–871.
- Oldfield, B. M., & Baron, S. (2000). Student perceptions of service quality in a UK university business and management faculty. *Quality Assurance in Education*, 8(2), 85–95.
- Osborne, S. P. (2010). *The new public governance? New perspectives on the theory and practice of public governance*. London: Routledge.
- Otley, D. T. (1999). Performance management: A framework for management control system research. *Management Accounting Research*, 10(4), 363–382.
- Pettersen, I. J. (2015). From metrics to knowledge? Quality assessment in higher education. *Financial Accountability & Management*, 31(1), 23–40.
- Pollitt, C., & Bouckaert, G. (2000). *Public management reform: A comparative analysis*. Oxford: Oxford University Press.
- Rabovsky, T. M. (2014). Using data to manage for performance at public universities. *Public Administration Review*, 74(2), 260–272.
- Rebora, G. (2003). La valutazione degli Atenei: alla ricerca della qualità. *Azienda Pubblica*, 5–6, 411–437.
- Rhodes, R. A. W. (1996). The new governance: Governing without government. *Political Studies*, 44(4), 652–667.
- Riccaboni, A. (2003). La valutazione integrata della didattica e della ricerca: il progetto VAI nell'Università di Siena. *Azienda Pubblica*, 5–6, 463–493.
- Schachter, H. L. (1995). Reinventing government or reinventing ourselves: Two models for improving government performance. *Public Administration Review*, 55(6), 530–537.
- Smith, G., Smith, A., & Clarke, A. (2007). Evaluating service quality in universities: a service department perspective. *Quality Assurance in Education*, 15(3), 334–351.
- Stoker, G. (2006). Public value management: A new narrative for networked governance? *American Review of Public Administration*, 36(1), 41–57.
- Talbot, C. (2005). Performance management. In E. Ferlie, L. E. Lynn, & C. Pollitt (Eds.), *The Oxford handbook of public management*. Oxford: Oxford University Press.
- Taylor, J. (2001). Improving performance indicator in higher education: The academics' perspective. *Journal of Further and Higher Education*, 25(3), 379–393.
- ter Bogt, H. J., & Scapens, R. W. (2012). Performance management in universities: Effects of the transition to more quantitative measurement systems. *European Accounting Review*, 21(3), 451–497.
- Turri, M. (2005). *La valutazione dell'università*. Milano: Guerini e Associati.
- Vakkuri, J., & Meklin, P. (2003). The impact of culture on the use of performance measurement information in the university setting. *Management Decision*, 41(8), 751–759.
- Van Dooren, W., Bouckaert, G., & Halligan, J. (2010). *Performance management in the public sector*. London: Routledge.

- Van Thiel, S., & Leeuw, F. L. (2002). The performance paradox in the public sector. *Public Performance & Management Review*, 25(3), 267–281.
- Waugh, R. F. (2002). Academic staff perceptions of administrative quality at universities. *Journal of Educational Administration*, 40(2), 172–188.
- Williams, I., & Shearer, H. (2011). Appraising public value: Past, present and futures. *Public Administration*, 89(4), 1367–1384.
- Yang, K., & Holzer, M. (2006). The performance-trust link: Implications for performance measurement. *Public Administration Review*, 66(1), 114–126.
- Zuccardi Merli, M., & Bonollo, E. (2013). I controlli manageriali nelle università. In L. Anselmi, F. Donato, L. Marinò, A. Pavan, & M. Zuccardi Merli (Eds.), *Il declino del sistema dei controlli manageriali nelle pubbliche amministrazioni*. Milano: Franco Angeli.

Chapter 17

Performance Management Uses, Outcome Measures and Mechanisms-Based Explanations. The Case of the Judicial Sector

Giancarlo Vecchi

Abstract This chapter analyses the different meanings of the concept of performance management utilisation, using empirical evidence from the Italian judiciary system. Managerial and evaluation literature on the ‘uses’ of performance information suggests that each use is sustained by specific causal mechanisms, i.e. strategies of actors that can explain—in specific contexts—the reasons for the success of the performance measurement. Empirical evidences for this view are derived from two case studies involving offices of the Italian judicial sector. The judicial sector has been selected because it is at the centre of an interesting debate over the value of the paradigm shift from output to outcome measures.

Keywords Justice sector · Performance management · Outcome indicators · Performance management use · Causal mechanisms

17.1 Performance Management Uses: Analysis of the Literature and the Research Questions

The chapter presents an analysis concerning the performance management uses in the public sector, with the aim to find evidences about how to foster the institutionalisation of these monitoring and evaluation processes. This analysis is developed considering the judicial sector, a field not often studied by the performance management literature and crossed by an interesting debate concerning the role of performance measures; in particular, the experiences of two Italian judicial offices are synthesized, considered good practices in the implementation of these instru-

G. Vecchi (✉)

Department of Management, Economics and Industrial Engineering - Policy Analysis Team,
Research Group: ReMAPS - Rethinking Management for Arts and Public Services,
Politecnico di Milano, Via Lambruschini 4/B (Campus Bovisa), Building 26B - Room 1.36,
20166 Milan, Italy
e-mail: giancarlo.vecchi@polimi.it

ments. The goal is to draw hypothesis about how to improve the utilisation of a performance management system, considering the mechanisms that can help leaders of public units in implementing it, overcoming organisational resistances.

First of all, a literature review about the classification of performance management uses is considered. Then, a synthetic introduction to the causal mechanisms theory is provided; this approach can help in developing hypothesis regarding the processes to improve the different usages of performance information. The following chapters present the debate on performance measurement in the judicial sector and in the Italian case; then a synthesis of two empirical case studies serves to test the proposed hypothesis. Finally, we draw conclusions about the contribution of this analysis to the research on performance management institutionalisation in the public sector.

Literature, with contributions from both performance and evaluation research, discusses often the problem of both the non-use (Pollitt 2013; Kettl 2016) and the different ways of performance measures utilisation. Considering our focus on the uses, it presents generally the following categorisation:

- (a) Instrumental use;
- (b) Process use;
- (c) Conceptual or enlightenment use;
- (d) Symbolic or legitimating use.

The *instrumental* use refers to the direct use of performance measures (or evaluation results) in making decisions about programmes. In such cases, managers learn from data to implement changes that are consistent with the original programme design or to ensure the likelihood of reaching the expected results (Vedung 2009).

Concerning the *process use*, the goal here is the development of usable knowledge through the interactions among the actors involved in the design, implementation and evaluation phases of a programme (Patton 1998; see also: Kirkhart 2000; Henry and Mark 2003; Hofstetter and Alkin 2003; Van de Walle and Van Dooren 2008, 2010).

The *conceptual* use focuses on the learning process resulting from the data analysis and evaluations, and can have as its object the programme's design, its implementation processes, and the outcomes. For example, organisational learning processes can be categorised as single-loop or double-loop (the former more devoted to efficiency, and the latter to effectiveness of programmes); more recently, a triple-loop learning process, focused on strategies and the values of an organisation, has also been proposed (Argyris and Schön 1978; Gilson et al. 2009). Weiss defined the conceptual use as *enlightenment*, to underline how the effects of learning are not always directly linked to the current policies but can indirectly influence policy decisions, for example when the results of the analysis inform in the middle period the problem setting phase, thus supporting the success of new frames to understand the issues at hand (Weiss 1998).

The last category in the usage taxonomy is the *symbolic use*, when data is used to support the status of a programme or an organisation administrator/leader. Here, literature underlines that data and evaluations could be used in two different ways. First, they can play a ‘substitute role’: an actor communicates that performance management is the first step to acknowledging a problem, while proper actions that should be taken are postponed or completely ignored. In the second way, data play a ‘legitimising role’: measures are selectively used to support an already held position or a decision made on a different base (Hofstetter and Alkin 2003, pp. 199–200). This role can also be used to sustain changes or to justify organisational isomorphism—the use of performance measurement to mimic the current mood of other organisations.

The literature on organisational and policy change identifies several external factors (e.g. focusing events, failure of policies, modification of public opinion, alterations in the socio-economic phenomena, etc.) as triggers of substantial changes in public structures. At the same time, it also highlights (more incremental) changes driven by the policy actors’ capabilities to learn from performance feedback (Kuipers et al. 2014). In the case of the Italian judiciary, external pressure from both the economic system and the users’ associations suggests that the instrumental use of performance feedback should be pursued. At the same time, weak managerial competences among magistrates¹ and staff, and internal opposition, indicate that conceptual and symbolic uses may be appropriate solutions; and the process use could be interpreted as an opportunity to improve internal and external cooperation. This gives rise to the hypothesis that organisational leaders, interested in the role of performance management systems to improve the quality and effectiveness of judicial services, will take the opportunity to use information and evaluations for a plurality of strategies, not only the instrumental one. A second hypothesis concerns the processes to implement the different uses of performance measures; the idea is that these processes can be conceptualised resorting to the causal mechanisms approach. To develop this last element, a discussion on the main characteristics of the approach is presented in the next paragraph.

17.2 Causal Mechanisms Fostering the Use of Performance Management Measures

Developing explanations by studying causal mechanisms is common to many social science disciplines. In policy analysis and public management, this approach is mainly used to explain the capacity of public programmes and reforms to reach the expected/observed outcomes. In particular, it attempts to explain the strategies that policy actors can adopt in designing and implementing change interventions.

¹The term ‘magistrate’ is used in the European meaning throughout the text, to indicate both judges and public prosecutors (see OECD 2013b).

Studies based on causal mechanisms analyse the black box that links a programme (e.g. a new organisational process) with the observed results. Researchers often view a programme primarily in terms of its effects, paying little attention to how the effects are produced. The mechanisms approach is a sort of ‘clear box’ analysis that aims ‘... to “unpack” the black box so that the inner components or logic of a programme can be inspected’ (Astbury and Leeuw 2010, p. 364).

Within this perspective, mechanisms are defined as factors that transmit ‘causal forces’ from an intervention to the outcomes, a sort of dynamic energy that explains the process through which a programme can modify, in a certain context, the problem situation (Beach and Pedersen 2013, p. 49).

When applied to public policies and organisations, this approach can explain the different ways in which change programmes are capable of modify actors’ ideas and behaviours. It is accepted that, given the role of specific contexts in the explanation of outcomes, the activation and effectiveness of a mechanism depend on the characteristics of the implementation environments. This conclusion does not prevent transferability strategies: mechanisms are conceptualised as theories that are ‘portable’ in the sense that they are building blocks for the adaptation of programmes to specific contexts and policy domains. As middle range theories (Merton 1968), they can be adapted through the research of ‘functional equivalents’ of their components. (Barzelay 2007; Astbury and Leeuw 2010; Pawson 2013).

In this chapter, we refer to this approach to determine the mechanisms that can foster the different uses of performance management, in particular when performance measures can help organisational leaders in introducing changes and innovations; and we hypothesise that each different performance management use can be improved by a specific set of mechanisms.

The managerial literature (Kelman 2005; Barzelay 2003, 2007; Barzelay and Jakobsen 2009) offers some useful, but not necessarily complete, indications. Considering the instrumental use, we can hypothesise the role of learning mechanisms that foster the capacity of organisations to deal with implementation problems. For example, ‘positive/negative feedbacks’ mechanism refers to the organisational learning based on the analysis of data provided by a performance management system; ‘sense-making’ regards the translation of data and information into definite programme modifications (Henry and Rog 1998); ‘naming and shaming’ is based on the disclosure of information to contrast non-compliant behaviours (Pawson and Tilley 1997; Astbury and Leeuw 2010); ‘focusing events’ are critical occurrences that can be used in organisations to improve the use of data and analysis, with the aim to prevent the return of conditions considered bad (Birkland 2007). The process utilisation can take advantage from mechanisms triggered to improve mutual learning to foster collaboration and avoiding conflicts; ‘repeated interactions’ regards the opportunity to use performance data in multiple interaction among different organisational actors to encourage consensus building (Kirkhart 2000); ‘reduction of cognitive dissonances’ is a mechanism triggered to facilitate the reception of innovations by reluctant actors, even with the help of performance data (especially outcome measures) (Kelman 2005); ‘integration’ is a mechanism that support the goal of cooperation among different units of an

Table 17.1 Types of performance management use and associated mechanisms

Utilisation types	Main mechanisms sustaining the use
Instrumental use	Positive/negative feedback (direct learning through use of evidences)
	Naming and shaming
	Sense-making
	Focusing events
Process use	Repeated interaction (mutual learning)
	Reduction of cognitive dissonances
	Integration processes
Conceptual/enlightenment use	Framing (foster a new vision)
	Focusing events
Symbolic/legitimation use	Perception of effectiveness
	Attribution of opportunity (policy entrepreneurship)

organisation or different structures (Pawson 2013). The conceptual/enlightenment use of performance management systems can be sustained through almost two mechanisms, when organisational leaders need to improve strategic shifts regarding how to deal with specific problems; the ‘framing’ mechanisms can be triggered to define a new way in formulating a problem and support a paradigm change, using the support of performance measures (Kahneman 2011); and the influence of focusing events can also be the starting point to use performance information and to support the relevance of a new strategy. Finally, the symbolic/legitimation use emerges when performance measures are collected and communicated to improve the reputation of actors and organisations; here the mechanism of ‘perception of effectiveness’ sustain this kind of utilisation through accountability processes; and the ‘attribution of opportunity’ mechanism explain why, in certain contexts, organisational actors consider performance management system as an incentive and use it as a way to improve their status (Barzelay 2007) (Table 17.1).

In the following paragraphs, we examine our hypothesis by analysing the development of performance management in the judicial sector. The next section introduces the debate on the shift to outcome measures in the justice sector of the OECD countries. Then, two cases—derived from recent practices in Italy—are presented and discussed. The aim is to explain the utilisation of performance management systems and the role of outcome measures in these experiences. Lastly, an analysis of the main mechanisms triggered to support the implementation and acceptability of a performance systems is proposed.

17.3 Performance Measures in the Judicial Sector

The role of the performance management systems in the judicial sector of OECD countries is a high debated topic, particularly the periodic collection of outcome measures and its impact on users and on the economic system. On the one hand,

citizens and companies seek prompt decisions, faster civil proceedings, more predictable judicial decisions and easy accessibility to justice services. On the other hand, magistrates believe that ‘justice’ is principally a public function, that cannot be considered (only) as a public service; therefore, qualitative inputs rooted in the magistrates’ professional knowledge should be considered as a priority, to safeguard (a) the autonomy and independence of the magistrates; and (b) the production of judicial decision based on high professional competence (avoiding the pressures of day-to-day public opinion and media).

Generally, the main performance management framework used in judicial offices is based on the input–output work flow: it considers the flow of cases from the filing phase, which is the incoming procedure, to the final phase, which is the resolution (using the available disposition alternatives). The performance measures derived from this set of information are: the quantity of pending cases at the beginning of the period (a given amount of time: a year, a month, etc.), an indicator of the level of workload coming from the previous periods; the incoming/reactivated cases (the cases filed during the period); the sum of both is the number of cases to be processed within the defined period; the number of disposed cases (i.e. resolved cases—decided as per judicial rites, dismissed, etc.), an indicator of the organisational productivity; and the total of active pending cases, or trials that remain in the ‘warehouse’ for the next period.

Since the 1990s, some OECD countries have developed more refined performance measures, oriented to the quality and outcome indicators of the judicial processes. They rely on an extended version of the input–output model, which considers the age of the active pending caseload (the number of days, cases have been pending or awaiting resolution) and time to disposition (the time elapsed from the date of filing to resolution, often compared with an agreed-upon case-processing time standard) (Ostrom and Hanson 2010). These indicators represent both efficiency characteristics and quality/effectiveness traits of ‘justice as public service’. For instance, the greater the length of trials (or of investigations), the longer the waiting time for the parties interested in the decision. Longer wait times, in turn, adversely affect economic and social affairs, both in civil and criminal fields. Moreover, they can determine the prescription of lawsuits and the end of criminal trials. Inordinately long delays represent de facto a denial of justice, and damage the citizens’ faith in the judiciary system. Another relevant outcome measure is *reversal rate*: the judicial decisions that are reviewed by a court of appeal (first or second instance). It denotes an important benchmark if analysed from an ongoing perspective, reflecting the quality of the judiciary policy choices.

The ‘justice as public service’ paradigm has been improved in the recent years by the international movements promoting public sector reforms in many OECD countries (OECD 2009, 2011, 2013a): for example, the new public management and the total quality analysis (Talbot 2010; Pollitt and Bouckaert 2011), the ‘public value’ paradigm (Moore 1995, 2013), the ‘New Public Governance’ movement (Osborne 2010; Christensen et al. 2014). In the past decades of the twentieth century, these reform movements attracted the attention of the judiciary system.

Accordingly, proposals and experiences of judicial use of performance management have been studied in many countries and by supranational institutions like

the European Commission (Fabri and Langbroek 2000; Langbroek 2005; Vigour 2006, 2015).

Over time, information on improvements needed in the performance management processes has become available, specifically in dimensions such as:

- Accessibility of services (for professional and non-professional users);
- Improvements in digitisation and communications programmes;
- Quality of interactions between users and magistrates/administrative staff, with the aim to improve the criteria of fairness, equality, respect, etc.

Efforts have also been made to link interventions on performance measurement and proposals focused on services' quality management. This leaves considerable room for the inclusion of internal process indicators and outcome measures such as:

- The satisfaction of customers and human resources;
- The various dimensions of accessibility;
- The trust and confidence of the public (citizens, companies, etc.) in the judiciary.

Judicial systems have also been persuaded to develop performance indicators and implement evaluation analysis because of the pressure imposed by scholars in economics and by international institutions. The quality and efficiency of the judiciary have been identified as a relevant factor for the functioning of markets and national (and regional) competitiveness (OECD 2013a, p. 8; World Bank 2011).

This literature offers many useful indicators to measure and compare the functionality of justice systems across different countries. The length of trials is, of course, the first factor suggested: a timely resolution of disputes indeed prevents firms from suffering undue expenses that may diminish their competitiveness. Other measures proposed include the following:

- The predictability of judicial decisions, as a way to guarantee the certainty of rules;
- The diffusion of alternative dispute resolution or simplified judicial procedures;
- The costs associated with the accessibility of judicial services, to avoid exclusion from judicial services;
- The uniformity and simplicity of procedures, with the aim to reduce the costs of compliance.

17.4 Performance Management in the Italian Judiciary

17.4.1 Introduction

Despite some relevant examples and a vast body of literature, the use of performance management has not witnessed sufficient diffusion in the Italian judicial offices. Currently, the Ministry of Justice is responsible for data mining: it monitors

main performance measures related to the functionality of judicial offices (e.g. Ministero della giustizia 2015). The single offices—generally—do not have any local performance measurement systems, and most of them lack the analytical competences needed to deal with performance management processes. They largely depend on data interpreted and disseminated by the Ministry of Justice and the Council for the Judiciary.

However, in the late 1990s, some of the main offices introduced practices of internal data interpretation. Presidents of courts and public prosecutors began to deal with issues of efficiency and quality of services, recognising the need for specific data, useful to represent the real organisation and to cover the processes of service production.

Following these initial experiences, over the last decade, a community of magistrates involved in the development of performance measurement systems has emerged, and some cases of successful implementation have come to light. The next paragraph discusses two such cases, that are considered good practices, followed by other offices, with reference to the modernisation processes implemented, and the efforts in the development of monitoring and accountability instruments (OECD 2013b; Politecnico di Milano 2015; Vecchi 2015; CSM 2016). The objective is to examine the uses of performance measures and the underlying mechanisms that helped the success of the implemented systems. Data, information and judgments are based on three sources: (a) an action research approach during six years, from 2009 to 2015, with an organisational development goal, implemented—through the *Innovagiustizia* ('Modernizing Justice') Project—by some units of the Politecnico di Milano, with the direct participation of the author in meetings and seminars connected with the elaboration of the performance management instruments (see Politecnico di Milano 2015); (b) interviews to the main actors involved in the analysed interventions (magistrates and staffs) of the two offices, collected by the author with the aim to reconstruct the change processes, actors and strategies (Pettigrew 2007; Vecchi 2013); (c) interviews to the main actors of the two offices, collected by an unit of the Italian Ministry of Public Administration, coordinated by the author, as part of an evaluation research with the goal to analyse the results of the 'Diffusion of Best Practices in the Italian Judicial System' National Project (Vecchi 2013, 2015; see Dipartimento della funzione pubblica (2015) and the bibliography for details).

17.4.2 Two Cases of Performance Management Use

The two cases analysed here are the Court of Milan and the Public Prosecutor's Office of Milan. Both are autonomous structures that act as the first jurisdictional level in the Italian judicial system. The Public Prosecutor Office manages the investigations concerning the criminal proceedings, while the Court is the judge of first instance. Considering the procedural flow of a criminal proceeding, the former starts the investigations and proposes a punishment to the Court judges, that can

reject it or start a trial that will finish with a first-instance sentence (that can be appealed). Both offices implemented performance management instruments, which involved capturing outcome measures and indicators.

The Court of Milan Case In 2011, this Court, one of the biggest judicial offices in Italy, started the implementation of an internal monitoring system, realised through an autonomous project and part of a broad modernisation programme based on the digitisation of work-processes. The system produced data regarding the input–output process (Tribunale di Milano—Comune di Milano 2011). The complete flow of cases was mapped, from the filing to the disposal phase. Two of the most significant achievements of the performance system were as follows:

- Data about cases were now available in real time for every judge, for every organisational unit and for the whole office;
- Data about cases were now available in a disaggregated way for more the main juridical subjects.

Some outcome indicators were regularly captured, especially the length of the trials. One such statistic was the number of cases continuing for over three years after their incoming day—trials that extend this time limit constitutes, under European regulations, a violation of the right to a reasonable duration of proceedings (beyond that time, an involved party can appeal for compensation on grounds of excessive length of a trial) (Table 17.2).

The Court also implemented other instruments to collect and disseminate outcome and quality data. This included user satisfaction surveys; for example, a survey conducted in 2013 (15 companies) and again in 2014 (2,029 companies), collected evaluations on items such as communication between the Court and users, accessibility of judicial services, costs of accessing those services, promptness of decisions and level of trust.

Table 17.2 Court of Milan—Example of monitoring data: civil sector

	2012/2013	2013/2014	2014/2015
Pending cases at the beginning of the period	117,501	120,503	116,143
Incoming cases	154,826	137,533	176,351
Decided cases	147,704	144,492	186,403
Pending cases at the end of the period	124,623	113,544	106,091
Simple clearance rate	95%	105%	106%
Full clearance rate	54%	56%	64%
Average time to define cases (with sentence)—in days	868	860	734
Average time to define cases (without sentence)—in days	414	441	441
Pending cases—more than 3 years old/total pending cases at the end of the period	–	13.1%	10.9%

Source Tribunale di Milano (2015)

Another instrument used for the diffusion of performance information was an annual report, voluntarily published since 2010 (see for example the last report: Tribunale di Milano 2015) that contained time series of performance indicators, some comparisons with other Italian judicial offices, the results of the user satisfaction survey and data on the progress of innovation projects.

The Public Prosecutor Office of Milan Case This judicial office shares with the Court of Milan the same relevant role in the Italian judicial system. It was continuously at the centre of a robust conflict against the political power for many years. In 1992, the magistrates of this institution started the ‘Clean Hands’ investigations against political corruption in Italy, which led to the end of the so called Italian First Republic; and in the following years, it managed many important investigations on political and economic cases.

In 2011, the chief of office launched a modernisation initiative, following the example of the Court of Milan and in cooperation with it. As a part of this initiative, the Public Prosecutor Office implemented the following performance management activities:

- A self-assessment exercise based on the Common Assessment Framework, and
- The elaboration and diffusion of an annual report, containing the main performance indicators and some outcome measures.

The self-assessment exercise (Procura di Milano 2012a, b), implemented between 2011 and 2012, led to a structured exchange of ideas and proposals between representatives of magistrates and the staff on one side, and among the different departments of the office, on the other side. When it was implemented, for the first time the divisions of the office exchanged information and judgements, through a formalised process, on organisational matters. During the meetings of the steering group, a set of performance indicators and a survey involving all the personnel—magistrates and staff (more than 350 people)—were implemented to support the assessment. The results of the process were presented to the office management (Chief and additional prosecutors) and were used to define interventions to improve the efficiency and quality of services (e.g. simplifying the procedure to visit an inmate and designing a process to define objectives and evaluate personnel). Moreover, the collection of reliable performance data was improved (e.g. data to monitor the wiretapping activities).

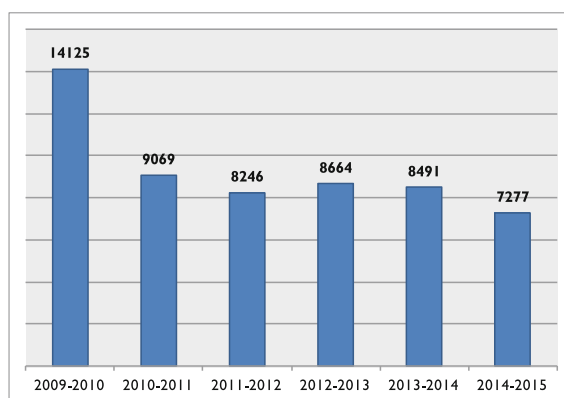
The publication of the annual report (since 2011) ensured the regular collection of a set of performance measures (see for example Procura di Milano 2015): reversal rate indexes (example: % of acquittals decided by judges—see Table 17.3); capacity to deal with crimes with strong social impact (e.g. pickpocketing, residential burglaries, sexual offences); length of investigations; number, duration and costs of phone tapping (see Table 17.4); quality of the services delivered; users’ satisfaction with front-office services; reliability of the information provided; waiting times; empathy.

Table 17.3 Public Prosecutor Office of Milan—Reversal Rate Measures: number and % of acquittals decided by the judges against the Public Prosecutor Office proposal (extract); number of precautionary measures rejected by the judges (extract)

Type of decision	Judicial year			
	2011/2012	2012/2013	2013/2014	2014/2015
Shorten trial sentence				
Acquittals	295 (22%)	416 (28%)	458 (29%)	473 (31%)
Conviction/mixed sentences	1044	1057	1115	1069
Precautionary measures				
Rejected by the judge	22.86%	25.68%	25.30%	30.00%

Source Procura di Milano (2015)

Table 17.4 Public Prosecutor Office of Milan—number of phone-tapping targets per judicial years (the objective was to reduce incrementally the number of targets, to avoid excesses and follow the recommendations of the Italian and European legislation)



Source Procura di Milano (2015)

17.5 Use of Performance Measures and Mechanisms Triggered

We found that both offices utilised the information provided by performance management instruments with some differences; in any case, the mechanisms triggered to promote the use belong to the categories presented at the beginning.

The *instrumental use* was evident in the direct utilisation of data to re-design procedures and work-processes. The main mechanism that sustained this type of utilisation was, as expected, the analysis of ‘positive/negative feedback’; it is usually at the base of all learning processes and reactions. In the Court of Milan the main evidence of this mechanism was the establishment of committees, at different levels, to monitor and discuss the key indicators regarding the progresses of projects

and the organisational performances. Other factors that confirmed the mechanism were changes in the management of work-processes, the training initiatives to bridge competency gaps and the reinforcement of innovation projects. In the Public Prosecutor Office, the same mechanism motivated the information collection through the self-assessment processes and the interventions for improving both the efficiency of internal procedures and the quality of services. Other mechanisms connected with the instrumental use were developed only at the Court of Milan. The ‘naming and shaming’ effect, operating through the diffusion of information about the state of the trials (through the so called ‘*Judges’ Console*’), was used informally to overcome the challenge of adopting hierarchical powers. The collection and diffusion of data thus represented an effective instrument to affect individual and organisational behaviours. Moreover, the instrumental use was supported in both offices by the ‘sense-making’ mechanism, triggered by the results of the surveys involving users and stakeholders. The opportunity to reflect on data and judgements from users helped to reinforce interventions based on the new services-oriented paradigm. Finally, also the ‘focusing event’ factor was relevant (e.g. the Milan Expo event), in showing the urgency of learning processes and immediate changes in both organisations.

The *process use* of performance information had in both offices a ‘multi-actors’ learning effect: a practice that depended on the mechanism of ‘repeated interactions’. The density of relations among the actors, constant interactions triggering learning processes (Dente 2014), and attitudes to reducing cognitive dissonances between traditional behaviours and the new vision (Kelman 2005) fostered integration practices (Greenhalgh et al. 2009). The two offices also presented some differences. The Court leveraged the performance instruments and measures to improve both internal and external relations. For instance, internal committees and the innovation unit were established to improve integration and mutual learning among different structures, in line with the collection and analysis of performance data. Moreover, the availability of performance measures helped strengthen relations between the Court and other external cooperative actors (e.g. the Milan Bar Association, the Municipality, etc.) through committees and groups meetings, like the ‘Justice Table of Milan’ (a committee composed by the main central and local public institutions). In those meetings performance reports were used as a means to share information and knowledge about service improvement for users and to develop ideas for solutions to organisational issues. The Public Procurement Office utilised performance management mainly internally. The collection of data about the quality of service delivery enabled a closer analysis of procedures characterised by the interaction between professional and non-professional users.

The *conceptual/enlightenment use* of performance information was also observed in both cases. In the Court and Public Prosecutor Offices, this use was mainly triggered by a ‘framing’ mechanism, with the goal to systematically sustain the concept of ‘justice as public service’, placing emphasis on outcome measures, for example the reduction of the length of trials and the improvement of quality of services. The measurement and evaluation activities, with their emphasis on the organisational weaknesses and the need for ICT development, were used to deploy an internal culture favourable to the modernisation paradigm. This translated not

only into a strategy to directly modify a single programme, but also to sustain a new vision of justice functions and to strengthen the ongoing changes. Evidence for this mechanism was derived from the formal documents of the office (e.g. the Annual reports and the formal planning documents) and from public speeches and interviews of the top officials during the whole period. Another mechanism triggered to support the conceptual use of outcome data was ‘focusing events’: the impending event of Milano Expo 2015 was constantly used as an example to underline the need for better results.

Finally, the diffusion of performance and outcome indicators enabled the *symbolic use* of information and knowledge, which covers developing partnerships and maintaining cooperation with other actors, or protecting the organisation and its leaders from criticism. The officials at both offices received support (including financial resources) from local and national institutions through the mechanism of ‘perception of effectiveness’ (a sort of ‘belief formation’ effect). By using performance measures to improve accountability processes, they built an image of virtuosity (even beyond their real results) useful to develop relations with local, national and supranational actors. The same symbolic use of performance data also was sustained by the ‘attribution of opportunity’ mechanism. Performance management processes were viewed by officials as a means to legitimise and improve their reputation as ‘modernisers’—a label that strengthened their leadership within the community of magistrates.

17.6 Performance Management Uses and Mechanisms: Lessons and Remarks

The two cases present experiences of performance measures utilisation that comes from one of the more complex sectors of the public domain. In facts, in the judicial sector the role of performance measures is still debated. So, we think that the lessons derived can be interesting for other sectors too, as a compass to improve the development of performance management systems.

The first lesson regards the plurality of the performance data uses. In other words, the success of the performance monitoring instruments derives not from one single type of use, only; but from the simultaneous activation of different utilisation processes. In general, the literature stresses the relevance of the instrumental use as the main indicator of a performance management effectiveness, that is a signal of non-use defeat. Here, we have to underline that the processual use plays a relevant role too. In the case of the justice sector, for example, the development of a stronger collaboration between the professional and the staff personnel is a premise to reach better organisational objectives; in that case, the discussion in group of performance results can help the common reflection concerning the critical factors that affect the

organisational work. Moreover, it is possible to conclude that in these cases the instrumental effectiveness of performance management practices depends on the capacity to implement even the processual utilisation, because this latter helps in driving the comprehension of organisational changes; in these empirical cases, for example, the adoption of the 'justice as service' principle through organisational solutions (as the introduction of procedures to apply a customer orientation).

Furthermore, these experiences, developed in a context of innovation programmes and under the thrust of a new organisational paradigm (deployment of ICT-based procedures and quality orientation), show the relevance of the enlightenment and, at the same time, the symbolic uses. The first has the aim to support in the middle period the affirmation of the new organisational ideas, the emphasis on an outcome view in designing and evaluating programmes (e.g. the impact of justice on the social and economic community); a longitudinal analysis of these interventions allows to observe the relevance of this type in shaping the organisational culture of the two offices. The remaining use (the symbolic one) plays a relevant role in the games among actors, supporting leaders' legitimation and the development of trust relations through the accountability function of the performance measures (in particular, the outcome indicators). In our two empirical cases, the assistance of the local institutions (with financial aid) to sustain the innovation projects has been certainly favoured by the periodic presentation of monitoring data regarding the progress of the activities. In this sense, the symbolic use should be analysed separating the positive contributions from the opportunistic behaviours, developing a more complete representation of this type. The mainstream of the current literature links it only to the misuse of a performance management system. There are opportunistic behaviours that fall into this category, of course; but it is worth to distinguish the symbolic uses that contribute to leaders' legitimation and to the construction of cooperative relations among organisations (Pollitt 2013).

The second lesson regards the development of the association between performance management utilisation and some causal mechanisms. The two empirical cases show that each type of performance management utilisation can be supported by specific mechanisms, triggered to improve the opportunity for the development and usefulness of performance measures in intra- and inter-organisational processes.

On the basis of the empirical evidence discussed here, we believe our hypothesis about the relations between performance management uses and supporting mechanisms is beneficial, especially in sustaining the paradigm shift from output to outcome measurements in the public sector. It can drive research on the success or failure of performance management systems, analysing the capacity to serve the different types of utilisation. Moreover, it can enhance learning from good practices, by highlighting the different ways to trigger the suggested mechanisms, ways that can be adopted in other sites to improve the success of a specific performance management use.

Sources of Data, Meeting and Interviews from 2010 to 2016

1. Innovagiustizia (modernizing justice) Project:

- (a) Public Prosecutor Office: four cycles to elaborate four Annual Reports (2011–2012–2013–2014), two monthly meetings during the period November–January of the years 2012–2013–2014–2015; the CAF self-evaluation process: from November 2011 to April 2012, two monthly meetings.
- (b) Court of Milan: four cycles to elaborate four Annual Reports (2010–2011–2012–2013): two monthly meetings during the period November–January of the years 2010–2011–2012–2013–2014; five meeting to the improvement of the ‘Judges’ Console’ during the period 2011–2013.

2. Semi-structured interviews:

Name	Collected by the author	Collected by an evaluation unit of the Italian Ministry of Public Administration, under the coordination of the author
Mrs. Livia Pomodoro, President of the Court of Milan during the period 2008–2015	22 November 2013	8 July 2013
Mr. Claudio Castelli, project leader of the modernization projects of the Courts of Milan during the period 2009–2005, and ex-director of the Ministry of Justice	6 February 2014; 11 September 2014; 26 September 2014	8–9 July 2013
Mr. Enrico Consolandi, Court of Milan, project leader for ICT development	1 March 2010; 7 July 2011; 1 March 2013; 21 November 2016	9 July 2013
Mr. Roberto Bichi, currently President of the Court of Milan, formerly Vice-president:	21 July 2015	9 July 2013
Mr. Nicola Stellato, director of the staff structure of the Court of Milan	–	9 July 2013
Mr. Edmondo Bruti Liberati, former Chief of the Public Prosecutor Office of Milan during the period 2010–2014	20 July 2015	10 July 2013
Mr. Michele del Medico, director of the staff structure of the Public Prosecutor Office of Milan	–	10 July 2013

(continued)

(continued)

Name	Collected by the author	Collected by an evaluation unit of the Italian Ministry of Public Administration, under the coordination of the author
Mr. Giovanni Xilo, consultant of the Court of Milan for the innovation projects	27 November 2013; 26 September 2014	–
Mr. Prof. Emilio BarTEZZAGHI, Politecnico di Milano, project leader of the innovagiustizia project	–	9 July 2013

References

- Argyris, C., & Schön, D. A. (1978). *Organizational learning: A theory of action perspective*. Reading, MA: Addison-Wesley Publisher.
- Astbury, B., & Leeuw, F. L. (2010). Unpacking black boxes: Mechanisms and theory building in evaluation. *American Journal of Evaluation*, 31(3), 363–381.
- Barzelay, M. (2003). Introduction the process dynamics of public management policymaking. *International Public Management Journal*, 6(3), 251–282.
- Barzelay, M. (2007). Learning from second-hand experience: Methodology for extrapolation-oriented case research. *Governance*, 20(3), 521–543.
- Barzelay, M., & Jakobsen, A. S. (2009). Theorizing implementation of public management policy reforms: A case study of strategic planning and programming in the European commission. *Governance*, 22(2), 319–334.
- Birkland, T. A. (2007). *Lessons of disasters. Policy change after catastrophic events*. Washington: Georgetown University Press.
- Beach, D., & Pedersen, R. B. (2013). *Process-tracing methods. Foundations and guidelines*. Ann Arbor, MI: University of Michigan Press.
- Christensen, T., Fimreite, A. L., & Lægreid, P. (2014). Joined-Up government for welfare administration. Reform in Norway. *Public Organization Review*, 14, 439–456.
- CSM—Consiglio Superiore della Magistratura (Council for the Judiciary) (2016). *Dalle buone prassi ai modelli. Una prima manualistica ricognitiva delle pratiche di organizzazione più diffuse negli uffici giudiziari italiani*. Roma: CSM.
- Dente, B. (2014). *Understanding policy decisions*. Heidelberg: Springer.
- Dipartimento della Funzione Pubblica (Italian Ministry of Public Administration) (2015). *Miglioramento performance per la Giustizia*. Retrieved March 26, 2016 <http://www.risorseperlagiustizia.it>.
- Fabri, M., & Langbroek, P. (Eds.). (2000). *The challenge of change for judicial systems. Developing a public administration perspective*. Amsterdam: IOS Press.
- Gilson Ch., Dunleavy P. and Tinkler J. (2009). *Organizational learning in government sector organizations: Literature review. A Report to the National Audit Office from LSE Public Policy Group*. London: London School of Economics. Retrieved March 26, 2016 from https://www.nao.org.uk/wp-content/uploads/2009/02/n0809129_Literature_review.pdf.
- Greenhalgh, T., Humphrey, C., Hughes, J., MacFarlane, F., Butler, C., & Pawson, R. (2009). How do you modernize a health service? A realist evaluation of whole-scale transformation in London. *The Milbank Quarterly*, 87(2), 391–416.

- Henry, G. T., & Mark, M. M. (2003). Beyond use: Understanding evaluation's influence on attitude and actions. *American Journal of Evaluation*, 23(3), 293–314.
- Henry, G. T., & Rog, D. J. (1998). A realist theory and analysis of utilization. *New Direction for Evaluation*, 78, 89–102.
- Hofstetter, C. H., & Alkin, M. C. (2003). Evaluation use revisited. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International handbook of educational evaluation* (pp. 197–222). Dordrecht: Kluwer.
- Kahneman, D. (2011). *Thinking fast and slow*. New York: Farrar, Straus and Giuroux.
- Kelman, S. (2005). *Unleashing change. A study of organizational renewal in government*. Washington: Brookings.
- Kettl, D. F. (2016). Making data speak: Lessons for using numbers for solving public policy puzzles. *Governance*, 29(4), 573–579.
- King, J. A., & Pechman, E. M. (1984). Pinning a wave to the shore: Conceptualizing evaluation use in school systems. *Educational Evaluation and Policy Analysis*, 6(3), 241–251.
- Kirkhart, K. E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation*, 88, 5–23.
- Kuipers, B. S., Higgs, M. J., Kickert, W. J. M., Tummers, L. G., Grandia, J., & Van der Voet, J. (2014). The management of change in public organisations: A literature review. *Public Administration*, 92(1), 1–20.
- Langbroek, P. (2005). Quality management concerning judges, judgements and court service. In M. Fabri, J.-P. Jean, P. Langbroek, & H. Pauliat (Eds.), *L'administration de la justice en Europe et l'évaluation de sa qualité*. Paris: Montchrestien.
- Merton, R. K. (1968). *Social theory and social structure*. New York: The Free Press.
- Ministero della giustizia (Ministry of Justice) (2015). Giustizia civile. Misurare la performance dei tribunali. Roma: Ministero della giustizia. Retrieved April 06, 2016 from https://www.giustizia.it/resources/cms/documents/Performance_tribunali_italiani_settore_civile.pdf.
- Moore M. H. (1995). *Creating public value. Strategic management in government*. Cambridge, MA: Harvard University Press.
- Moore, M. H. (2013). *Recognizing public value*. Cambridge, MA: Harvard University Press.
- OECD (2009). *Measuring government activity*. Edited by Christopher Pollit, Geert Bouckaert and Wouter van Dooren. Paris: OECD.
- OECD (2011). *Economics of civil justice: An explorative study*. Paper prepared by Bruno Deffains. Paris: OECD.
- OECD (2013a). *Judicial performance and its determinants: A cross-country perspective. A going for growth report* (OECD Economic Policy Papers N 5, prepared by Giuliana Palumbo, Giulia Giupponi, Luca Nunziata, Juan Mora-Sanguinetti). Paris: OECD.
- OECD (2013b). Italy: The digitisation of the court of Milan: An opportunity for systemic innovation in the delivery of civil justice services to citizens. In *Innovation for better public services. 48th Session of the Public Governance Committee* (pp. 62–80). Paris: OECD. Retrieved May 10, 2016 from [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=GOV/PGC\(2013\)3/REV1&docLanguage=En](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=GOV/PGC(2013)3/REV1&docLanguage=En).
- Osborne, S. (Ed.). (2010). *The new public governance? Emerging perspectives in theory and practice of public governance*. Abington: Routledge.
- Ostrom, B., & Hanson, R. (2010). *High performance framework. Achieving high performance: A framework for courts*. Newport: National Centre for State Courts.
- Patton, M. Q. (1998). Discovering process use. *Evaluation*, 4(2), 225–233.
- Pawson, R. (2013). *The science of evaluation. A realist manifesto*. London: Sage.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London: Sage.
- Pettigrew, A. M. (2007). What is a processual analysis? *Scandinavian Journal of Management*, 13(4), 337–348.
- Politecnico di Milano (2015). *Innovating the justice system. Innovagiustizia: The Lombard Experience of the Project Diffusion of Best Practices in the Italian Judicial Offices 2010–2013*. (Synthesis of the project financed by European Commission/FSE Programme, Italian Ministry of Justice, Italian Ministry of Public Administration, Italian Ministry of Labour, and Lombardia

- Regional Government. Retrieved April 02, 2016 from http://media.wix.com/ugd/3b738c_a485aac562de4fec96374316cac366f1.pdf.
- Pollitt, C. (2013). The logic of performance management. *Evaluation*, 19(4), 346–363.
- Pollitt, C., & Bouckaert, G. (2011). *Public management reform: A comparative analysis—new public management, governance, and the Neo-Weberian State* (3rd ed.). Oxford: Oxford University Press.
- Procura di Milano (2012a). Analisi delle risposte al Questionario CAF giustizia. Milano: Procura di Milano (unpublished internal document).
- Procura di Milano (2012b). Sintesi dell'autovalutazione CAF. Milano: Procura di Milano (unpublished internal document).
- Procura di Milano (2015). *Bilancio di responsabilità sociale 2014/15*. Milano: Procura della Repubblica di Milano. Retrieved April 02, 2016 from <http://www.procura.milano.giustizia.it/bilancio-sociale.html>.
- Talbot, C. (2010). *Theories of performance. Organizational and service improvement in the public domain*. Oxford: Oxford University Press.
- Tribunale di Milano—Comune di Milano (2011). *Controllo di gestione e Gestione allarmi*. Specifiche funzionali. Milano: Tribunale di Milano (Internal document).
- Tribunale di Milano (2015). *Bilancio di responsabilità sociale 2015. A misura di cittadino*. Milano: Tribunale di Milano. Retrieved April 12, 2016 from https://www.tribunale.milano.it/index.phtml?Id_VMenu=241.
- Van de Walle, S., & Van Dooren, W. (2008). Introduction: Using public sector performance information. In Wouter Van Dooren & Steven Van de Walle (Eds.), *Performance information in the public sector: How it is used* (pp. 1–8). Houndmills: Palgrave.
- Van de Walle, S., & Van Dooren, W. (2010). How is information used to improve performance in the public sector? Exploring the dynamics of performance information. In K. Walshe, G. Harvey, & P. Jas (Eds.), *Connecting knowledge and performance in public services: From knowing to doing* (pp. 33–45). Oxford: Oxford University Press.
- Vecchi, G. (2013). Systemic or incremental path of reform? The modernization of the judicial system in Italy. *International Journal for Court Administration*, 5(1), 64–87.
- Vecchi G. (2015). Progetto Diffusione delle best practices negli uffici giudiziari italiani. Valutazione esplorativa di 5 casi positive. *Paper for the Italian Ministry of Public Administration, Project Monitoring the Performance of Justice*. Roma: Dipartimento della funzione pubblica.
- Vedung, E. (2009). *Public policy and program evaluation*. New Brunswick: Transaction Publishers.
- Vigour, C. (2006). Justice: l'introduction d'une rationalité managériale comme euphémisation des enjeux politiques. *Droit et société*, 63–64, 425–455.
- Vigour C. (2015). Current transformations to the judicial system in Europe: A new model of justice in the making. *Paper presented at the International Conference on Public Policy*, Milan 1–4 July 2015.
- Weiss, C. (1998). Have we learned anything new about the use of evaluation? *American Journal of Evaluation*, 19(1), 21–33.
- World Bank (2011). *Improving the Performance of Justice Institutions. Recent Experiences from Selected OECD Countries Relevant for Latin America*. (Paper prepared by Klaus Decker, Christian Möhlen, David F. Varela). Washington: The World Bank.

Chapter 18

Performance Management and Evaluation of Large-Scale Events in a Multistakeholder Engagement Perspective: The Case of the Extraordinary Jubilee of Mercy

Marco Meneguzzo, Gloria Fiorani and Rocco Frondizi

Abstract Large-scale events, which are characterized by a dynamic complexity involving different national public administrations and institutional levels, require a special attention in selecting adequate project management and event management systems, and in designing and implementing multistakeholder management and engagement techniques. They also require the adoption of innovative managerial tools, as well as the introduction of performance evaluation systems linked to an inter-institutional cooperation and collaboration setting (performance management at macro level). Using a multidisciplinary approach (public management and governance, political science, administrative science), this chapter analyzes the case of the Extraordinary Jubilee of Mercy, focusing on its outcomes in terms of inter-institutional and collaborative governance, cross-sector collaboration and joined-up government.

Keywords Performance evaluation · Event management · Network governance · Multistakeholder engagement · Inter-institutional cooperation

M. Meneguzzo (✉)

Public and Nonprofit Management, University of Rome Tor Vergata, Rome, Italy
e-mail: meneguzzo@economia.uniroma2.it

M. Meneguzzo

Università della Svizzera Italiana, Lugano, Switzerland

G. Fiorani · R. Frondizi

Public Management, University of Rome Tor Vergata, Rome, Italy
e-mail: fiorani@economia.uniroma2.it

R. Frondizi

e-mail: rocco.fronidizi@uniroma2.it

© Springer International Publishing AG 2018

E. Borgonovi et al. (eds.), *Outcome-Based Performance Management in the Public Sector*, System Dynamics for Performance Management 2, DOI 10.1007/978-3-319-57018-1_18

18.1 Introduction

Planning, implementing, and financing large-scale event in different domains (leisure and sports, such as the Olympic Games; arts and culture, such as major film festivals—e.g., Berlin, Cannes, Venice; or economic, such as International Expos—e.g., Expo Milan 2015) is an important challenge for national and international institutions, academics, and practitioners, for many reasons. Large-scale events cannot be clustered within the boundaries of a single public administration or agency, and they are characterized by a dynamic complexity involving different national public administrations and institutional levels (central/federal; regional; metropolitan/local), as well as a wide range of stakeholders and actors (public, private for-profit, and private nonprofit).

These events require a special attention in selecting adequate project management and event management systems, in designing and implementing multistakeholder management and engagement techniques, in adopting innovative managerial tools, such as public policies implementation scenarios, evaluability assessments, simulation models based on system thinking, system dynamics, and social network analysis. Another key decision is related to the evaluation systems for the economic, social, and environmental impacts generated by such macro events, as well as the design of methods and systems for evaluating the performance of the different public administrations and agencies involved in the different stages of policy formulation and policy implementation.

The topics of performance management and performance evaluation of public intervention policies require an enrichment and an integration of Bouckaert and Halligan's fundamental contribution on performance management based on the analysis of national case studies, on the variables of width and depth, and on the four models of performance management (Bouckaert and Halligan 2008).

In this perspective, we can recall Bouckaert's observations on performance management and hierarchies, markets, and networks, which he presented at the ASPA-EGPA Transatlantic Dialogue held in 2014 in Lugano (Switzerland) (Bouckaert 2014), as well as a recent interesting contribution on the role of the various institutional levels (Kuhlmann and Wayenberg 2016).

Interesting theoretical inputs can also be obtained from the analysis of the case study of the Extraordinary Jubilee of Mercy, co-designed and co-organized by two national governments (Italy and the Vatican State) and involving the Catholic Church as a recognized universal international institution and which can be linked to the model of governance performance (Bouckaert and Halligan 2008).

Our contribution starts from the analysis of this event, in which the authors have been directly involved as participants to the inter-institutional task force in charge of managing the event (led by the Technical Secretariat for the Jubilee).

This chapter aims to define the basics of a theoretical framework useful to understand cross-sector collaboration, joined-up government, inter-institutional, and collaborative governance. Given this interdisciplinary perspective, the literature review will link different disciplinary approaches and research streams, such as the

Table 18.1 Theoretical framework and research questions. *Source* own elaboration

Theoretical framework		
<i>Stakeholder management Network governance and management</i>	<i>Public policy implementation</i>	<i>Event management and project management</i>
RQ1: Is it possible to adopt stakeholder mapping tools in the analysis of inter-institutional public networks?	RQ2: Which public policy implementation approaches are the most appropriate for the management of highly complex events and inter-institutional projects?	RQ3: Were project management tools effective in the case of the Extraordinary Jubilee of Mercy?

evaluation of public policy implementation (political sciences, administrative sciences), project and event management, network management and governance (organizational sciences, sociology and management), stakeholder management (management and sociology), with the ambition of identifying possible synergies between them, in a cross-disciplinary perspective. Table 18.1 summarizes the theoretical approaches and the research questions of this work.

This chapter is organized as follows. In Sect. 2 we present a theoretical framework on the topics of policy implementation, stakeholder management and network governance, and management. Section 3 is dedicated to an in-depth examination of the existing literature on event management, with a focus on the project management approach in the design, implementation, and evaluation of large-scale events. In Sect. 4 this approach is adopted to analyze the case study of the Extraordinary Jubilee of Mercy. Finally, conclusions are presented in Sect. 5.

18.2 Theoretical Framework: Stakeholder Management, Network Governance and Management, Public Policy Implementation

Different disciplinary approaches could be useful for analyzing the design and the implementation of the event ‘Extraordinary Jubilee of Mercy’.

The first is linked to research streams on New Public Management (NPM), Public Governance (PG), and New Public Governance (NPG) (Pollitt and Bouckaert 2011), with the last one, in our opinion, being more adequate when studying large-scale events. At the international level, the NPG has been studied by several authors, with some relevant contributions based on the systematization of different national and functional/sectorial case studies (Bovaird and Loffler 2003; Osborne 2010).

At the Italian level, literature reviews on public governance, as well as national projects coordinated by the Italian Ministry of Public Administration and by the national task force on innovation in the public sector (in particular, the Governance

project launched by the Department of Public Administration in the period 2005–2007), have identified three main areas. These are external governance, regarding the system of relationships between local administrations and private stakeholders; inter-institutional governance, referring to the agreements and interactions between administrations, agencies, and government-owned firms; and internal governance, concerning single public administrations (Meneguzzo 1995; Cepiku 2005).

Moreover, our theoretical framework also includes studies related to the organizational and network management sciences, in particular to the coordination mechanisms in order to understand the effectiveness of the network in terms of guidance and building stakeholder commitment (Mandell 1999; Klijn 2010).

Prominent international scholars have suggested that, in complex and changing times, in fragmented societies, participative approaches (Agranoff and McGuire 2001; Milward and Provan 2003) should help accomplish outcomes in a more effective way. In their researches, complexity is referred to as the increased connectivity among organizations (public, private, and nonprofit actors) at different levels (local, national, and supranational), interacting in an external and multidimensional environment (economic, social, and cultural).

There is a growing need for contemporary governments to look at innovative methods to deal with complex and social problems, the so-called ‘wicked issues’ (Clarke and Stewart 1997), and networks seem to be at the forefront of this change (Mandell 1999); thus the engagement of public actors is considered the rule rather than the exception (Ferlie et al. 2005).

As a consequence, in the last stage of public governance, the shift to network management (Kickert et al. 1997) happened naturally according to the societal requirement going beyond the traditional models, which were inadequate to understand the features of the external environment, considered complex, dynamic, and diverse (Eljassen and Kooiman 1993). As stated by Isett et al., the study of networks in public administration literature has rapidly evolved, identifying three mainstreams that emphasize exactly how networks in the public sector are nowadays used as a mechanism to encourage collaborations (Isett et al. 2011). Shortly, these are (i) policy networks, as a set of public agencies, private sector, and nonprofit organizations with a common interest, looking for decisions and pursuing political agendas in a specific area of policy; (ii) collaborative networks, as pools of government agencies, nonprofit, and for-profit organizations, working to ensure citizens’ satisfaction in providing public services and goods; and (iii) governance networks, where collective actions of several actors are required to find solutions to public issues which cannot be solved by one individual.

As stated by Pollitt and Hupe, “networks are envisaged as the more adequate way to make sense of contemporary complexity”, and problems are identified in the lack of knowledge, the involvement of many interdependent players and various decision-making arenas (Pollitt and Hupe 2011). Thus, networks are considered multi-organizational arrangements in managing problems; they overturn traditional governing structures through formal and informal ties, characterized by reciprocity, mutual interdependencies, and various individuals and organizational actors involved in the formulation, implementation, and delivery of public services.

In this sense, public networks are recognized by scholars (e.g., Milward and Provan 2003), as well as by practitioners and even policy-makers, as a model of governance characterized by quality, flexibility, and innovativeness in improving coordination in order to increase efficiency and effectiveness (e.g., in the arenas of health and social care, local development, education, security, and culture). Network management investigates the way networks are governed and managed and, in particular, which are the mechanisms to promote and build stakeholders' commitment to the final achievement (Provan and Kenis 2008).

Another two research paradigms are very useful in interpreting and describing the Extraordinary Jubilee of Mercy as a complex event.

The first is related to stakeholder theory (Freeman 1984; Freeman et al. 2010) and more specifically to the activity of stakeholder mapping, as well as to the other two key activities represented by stakeholder management and stakeholder engagement.

The second is linked to the stream of analysis of the implementation of public policies, which represents an important research field at the international level in the areas of public administration, administrative sciences, and public policy (for a preliminary summary in the Italian context, see: Dente 1982). The analyses of policy and program implementation, which can be extended to complex events, can be divided into those based on top-down approaches, those based on bottom-up approaches, and those based on syncretic approaches. For what concerns top-down approaches, we can quote Pressman and Wildavsky (1973) and Wildavsky (1981) on the complexity of joint action and on implementation delays, and Bardach (1977) on the identification of different games in the implementation of public programs. For what concerns syncretic approaches, we should recall other US (Nakamura and Smallwood 1980) and German scholars (Hanf and Scharpf 1978).

Implementation analyses were used to evaluate programs and policies for the management of structural funds, starting from the MEANS 1999 program (Monnier 1999), and they were studied by Italian public management scholars (Mussari 1999; Rebora 1989; Meneguzzo and Del Vecchio 2000). Several important and complex events in Italy were analyzed using the approach of public policy implementation, especially by administrative sciences and urban planning scholars, starting from the 1990 FIFA World Cup, the '*Colombiadi*' held in Genoa in 1992, the 1997 Turin Winter Olympics, the 2009 Swimming World Championship in Rome and Expo 2015 in Milan.

18.3 Theoretical Framework: Event Management and Project Management

The way and approach by which project management supports the creation, development, and execution of large-scale events is commonly known as event management. As Thomas et al. point out, event management "comprises the

coordination of all of the tasks and activities necessary for the execution of an event regarding its strategy, planning, implementation, and control, based on the principles of event marketing and the methods of project management” (Thomas et al. 2008). Starting from some pioneering studies (Getz 1991; Hall 1992), event management has been rapidly rising and recognized as a profession and a formal discipline, reflecting the “need within all societies for the professional management of events in the private, public and not-for-profit sectors” (Getz 2005).

18.3.1 Definitions and Classifications of Events

The first step to understand the meaning and scope of event management is the definition of the event itself. Etymologically, the term event derives from the Latin events, which directly translates as “occurrence, issue”. Today, events form an integral part of all societies (Ferdinand and Kitchin 2012) and the recognition, by practitioners and researchers, of their role as an ‘industry’ has gained importance in recent times, also because of the quantitative and qualitative growth forecasts for their market (Thomas et al. 2008). In the academic literature, many authors from different disciplines have discussed the meaning and definition of events and related concepts, failing to reach a unanimous consensus on standardized terms, definitions, or categories.

According to Shone and Perry, an event is “that phenomenon arising from those non-routine occasions which have leisure, cultural, personal or organizational objectives set apart from the normal activity of daily life, whose purpose is to enlighten, celebrate, entertain or challenge the experience of a group of a people” (Shone and Parry 2004).

Getz defines events as “temporary occurrences, either planned or unplanned” (Getz 1991).

In order to emphasize the difference between unplanned and planned events, the term ‘event’ is preceded by the word ‘special’. A special event could be a “one-time or infrequently occurring event outside normal programs or activities of the sponsoring or organizing body”, as well as “an opportunity for a leisure, social or cultural experience outside the normal range of choices or beyond everyday experience” (Getz 2005).

All the above definitions stress the unique and temporary nature of events, a feature also highlighted in the definition by Silvers, who considers an event as “the gathering of people at a specified time and place for the purpose of celebration, commemoration, communication, education, reunion and/or leisure” (Silvers 2008).

Other authors, such as Jago and Shaw, focus on the tourism context, defining a special event as “a one-time or infrequently occurring event of limited duration that provides the consumer with a leisure and social opportunity beyond everyday experience”, often held “to raise the profile, image or awareness of a region” (Jago and Shaw 1998).

Summarizing these definitions, we can define an event as a unique or infrequently occurring happening that occurs at a given place and time outside normal activities for individuals or groups of people.

Concerning typologies, events are often classified according to different criteria. In terms of size, it is possible to identify four different types of events (Bowdin et al. 2006; Jackson 2013): local or community events, which are generally small and linked to a particular geography; major events, which attract significant numbers of visitors and media coverage; hallmark events, which are not automatically bigger than major events, but are synonymous with a particular place; and mega events, which are global and can influence the host country's economy.

In terms of form, while Bowdin et al. identify only three types of events—cultural, sports, and business events (Bowdin et al. 2006)—Getz (2005) distinguishes ten categories—cultural celebrations, religious events, political and state events, arts and entertainment, business and trade events, education and scientific events, sports events, recreational events, private events, and events at the margin. Finally, Raj and Musgrave (2009) differentiate between the following forms: religious events, cultural events, musical events, sporting events, personal and private events, political and governmental events, commercial and business events, corporate events, special events, and leisure events (Jackson 2013).

18.3.2 Project Management and Event Management

Until the late 1990s, the management of events was implemented in patchwork and disparate ways (O'Toole 2000). However, since the related environment had become more complex and events had started to be considered a key success factor in many sectors, the need emerged for a systematic, standardized, and accountable approach to planning and control. Project management could provide a solution to these problems.

Starting from the fundamentals of project management, a group of academics and practitioners has developed a common event management framework known as the 'Event Management Body of Knowledge' (EMBOK).

This model includes a definition of the phases, processes, and core values, as well as the knowledge domains to which these are applied. The EMBOK can be regarded as "an integrated, sequential, and iterative system associated with each element of each class of each domain at each phase in the event management process, with tools and techniques that may be used for each" (Silvers et al. 2006). In particular, the Phases (including initiation, planning, implementation, event, and closure, terms derived from traditional project management terminology) illustrate the sequential nature of event management, emphasizing the criticality of time in any event project. The Processes illustrate both a sequential and iterative system that promotes a comprehensive course of action and a dynamic approach to the changing nature of events. The Core Values (continuous improvement, creativity, ethics, integration, and strategic thinking) specify those principles that must

influence all decisions regarding every element, phase, and process of an event, to ensure these decisions facilitate successful and sustainable outcomes. Finally, the Domains (administration, design, marketing, operations, and risk) illustrate the full scope of activities and functions within event management. It should be noted that, unlike a pure business model which emphasizes outputs, “a major use for the EMBOK is in the recognition of event management as a process” (Silvers et al. 2006).

Bowdin et al., following the traditional principles of project management, identify five phases and ten knowledge areas to describe the processes involved in event management. The phases are initiation, planning, implementation, event, and shutdown, while the knowledge areas are scope, marketing, finance, time management, design, risk management, procurement, human resources, stakeholder management, and communication (Bowdin et al. 2011).

According to O’Toole and Mikolaitis, project management and event management strategies bring the following advantages: establishing a systematic approach to all events; depersonalizing the event; facilitating clear communication; conforming to the methodology used by other departments; ensuring accountability; increasing the visibility of event planning; facilitating training; developing transferable skills; and establishing a diverse body of knowledge (O’Toole and Mikolaitis 2002).

However, although project management is today a commonly accepted management framework for events, there are still limitations. As highlighted by Bowdin et al., project management often lacks the flexibility required by special events; other limitations include the contribution of volunteers, which is difficult to quantify, as well as the uncertainty related to some aspects of the management of an event, such as the number of stakeholders or the number of tickets sold (Bowdin et al. 2011).

18.4 The Case of the Extraordinary Jubilee of Mercy

The 2016 Holy Year is called ‘extraordinary’ not only because it did not fall on the 25 years canonic period from the previous, but also because it was the first ‘thematic’ Jubilee, dedicated to mercy, and the first to be ‘widespread’. With an unprecedented decision, Pope Francis chose to carry it out in all the cathedrals of the world, with each Diocese being able to open its Holy Door. After the first Holy Door, opened in Bangui (Central African Republic) in November 2015, more than 500 ‘Doors of Mercy’ were opened during the Holy Year in 91 countries. This Jubilee was also modern, digital, and built on a network basis. The 2016 Holy Year was, in fact, the first in the age of social networks, tablets, and smartphones, in the era of the continuous connection to the Internet.

The Extraordinary Jubilee of Mercy, called for by the Pope on April 11, 2015, started on December 8, 2015 and ended on November 20, 2016. More than 21 million people arrived in Rome in 2016, from 36 different countries. Such

exceptional amount of visitors and notable personalities has had important implications for the general system of hospitality, transportation, infrastructures, and for public security.

The Extraordinary Jubilee of Mercy was a complex event, because it required planning and managing in line with the events organized in its framework, which showed different levels of complexity based on the number of participants and on the nature of risks. 239 days were labeled as ‘ordinary’ (the ‘White Days’, without events in program), 49 as ‘special days—level 1’, 34 as ‘special days—level 2’, 16 as ‘special days—level 3’, and 11 as ‘great event—level 4’.

For each different typology, the Jubilee Management Room (*Sala Gestione Giubileo—SGG*) defined coherent interventions: ‘Ordinary’ for White Days, ‘Intermediate’ for Level 1 happenings, ‘Event’ for Level 2, 3 and 4 happenings. We should note that the SGG operated in the ‘Event’ mode in 55% of the ‘event days’ (61 days out of 110).

The strategic management of the event was assigned by the Italian and the Vatican State governments to the Prefect of Rome. The Prefect was in charge of defining the planning framework for managing the event and of ensuring that the whole information and communication systems worked.

A specific task force (implementation unit) was also activated as a way to facilitate the cooperation and collaboration between the different public administrations involved in the event. The ‘Technical Secretariat for the Jubilee’ (Fig. 18.1) included 10 experts representing public administrations and territorial institutions (Ministry of the Interior, Lazio Region, Municipality of Rome), as well as delegates from Civil Protection, Carabinieri Corps, and Fire Corps.

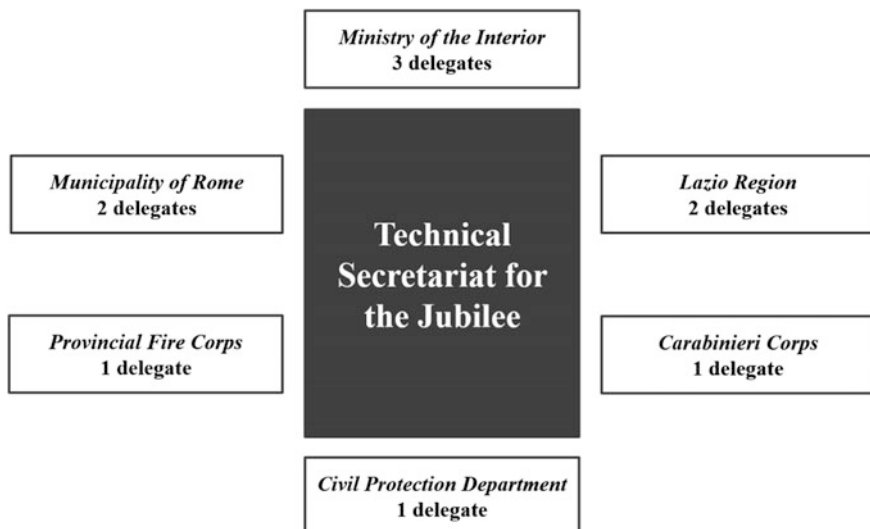


Fig. 18.1 The Technical Secretariat for the Jubilee. *Source* own elaboration

Under the management of the Technical Secretariat, the Prefect created eight thematic working groups, with the task of planning for the different areas, and invited all interested institutions to participate (Fig. 18.2).

Once the stakeholder mapping and thematic clustering were in place, an in-depth analysis of the expectations, needs, and interests of all stakeholders was necessary. All parties involved have been proactive and accepted to share their resources, their information, and competencies: this led to the creation of a new concept of large-scale event, which capitalizes on the synergies between all institutions involved, both at the intra-institutional and at the inter-institutional and intergovernmental levels, and between the eight working groups.

As we can see in Fig. 18.2, the key pillars of the governance model are on the one hand, the Prefecture of Rome and the Technical Secretariat for the Jubilee, the Municipality of Rome, Lazio Region and the Metropolitan City of Rome, and on the other hand, the eight thematic working groups, where ‘safety and security’ and ‘Italy-Vatican State intergovernmental relations’ were crucial in the management of the event.

In order to define a performance management system, what we find especially interesting is not only the work done in the single groups, which led to specific results, but also the relations, shown in Fig. 18.2, between the different groups, which require the use of performance management and measurement systems for the networks of public administrations which originated.

For what concerns internal governance, the working groups met regularly (every 3 weeks/ monthly) to define the project management systems for the event. In each meeting, the Technical Secretariat for the Jubilee drafted a report indicating

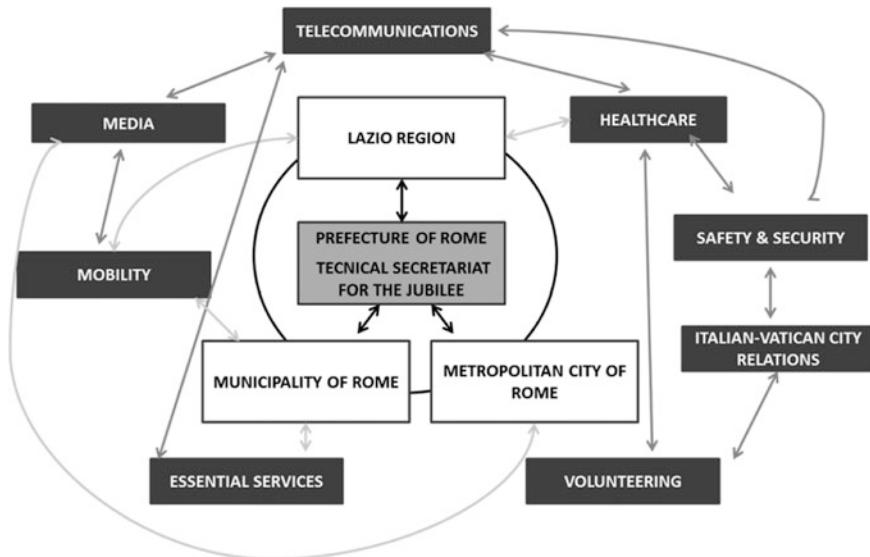


Fig. 18.2 Inter-institutional synergies. *Source* own elaboration

participants (at least one for each institution/public administration/company and the representative for the Technical Secretariat), issues to be solved and possible solutions. The Technical Secretariat and the eight working groups performed strategic and operational activities (Table 18.2 indicates the main ones, their outputs, and the actors involved in the working groups). The planning and setup stage, which culminated with the adoption by the Prefect of Rome of the “document describing the integrated planning framework for the whole management of the Jubilee” (December 7, 2015), was followed, during the whole Jubilee year, by an

Table 18.2 Main activities, actors, and outputs of the working groups. *Source* own elaboration

Working groups	Main activities	Main actors involved	Main outputs
1. Media	Management and coordination of the <i>Infomobility</i> service; identification of the authorities in charge in case of emergencies; management of tools (official websites, apps and social media dedicated to the Jubilee); coordination of information flows; management of brand and image ('Rome for the Jubilee' logo)	Technical Secretariat for the Jubilee; Municipality of Rome; Metropolitan City of Rome; Lazio Region; Ministry of Transportation; Ministry of Economic Development; Local Police; Civil Protection; Fire Corps; ATAC; Civitavecchia Port Authority; Ferrovie dello Stato; COTRAL; Grandi Stazioni	Communication Plan; implementation of a Media Center; creation of the <i>Pilgrim Help Desk</i> , to assist pilgrims and citizens and avoid frauds
2. Volunteering	Definition of the volunteers' profiles, activities and presence; management of the relationships between various volunteering associations; planning the collaboration between the Civil Protection and public services providers	Technical Secretariat for the Jubilee; Municipality of Rome; Lazio Region; Civil Protection; Local Police; Trenitalia; Ferrovie dello Stato; Grandi Stazioni; ATAC	Volunteers performed the following tasks: information and reception of pilgrims; first aid activities; support to Local Police for the management of the flows of pilgrims
3. Safety and security	Definition of an integrated planning framework for the management of the event; coordination of all the safety and security operators; management of critical situations and implementation of corrective actions	Technical Secretariat for the Jubilee; Municipality of Rome; Civil Protection; Italian National Police; Local Police; Fire Corps; Agency for the Mobility; ARES—118	Safety Plan, in accordance with the integrated planning framework; implementation of the Jubilee Management Room, in collaboration with Local Police; implementation of a shared cartographic system, in collaboration with the Technological Innovation Department of the Municipality of Rome and SOGEL; implementation of <i>SIGIS</i> , an environmental monitoring system, owned by the Provincial Fire Corps of Rome, which can detect anomalous substances in the air with an action range of up to 5 km

(continued)

Table 18.2 (continued)

Working groups	Main activities	Main actors involved	Main outputs
4. Telecommunications	Management and coordination of communication systems; use of the integrated communication system <i>Radio TETRA</i> , provided by the Local Police of Rome, which allowed to ensure simultaneous radio communications between all operative forces at the events	Technical Secretariat for the Jubilee; Ministry of Economic Development; Municipality of Rome; Civil Protection; Italian National Police; Local Police; Telecom Italia; Wind; Vodafone; Fastweb	Optimization of communication systems, especially for what concerns wifi spots and signal
5. Essential services	Planning and programming essential interventions and managing the urgent ones; definition of the relationships between the Jubilee Management Room and the main telecommunications and public services/utilities companies (energy, gas, water)	Technical Secretariat for the Jubilee; Ministry of Economic Development; Municipality of Rome; ACEA; AMA; ATAC; Italgas; Terna; Telecom Italia; Wind; Vodafone; Fastweb	Essential Services Plan, to plan and manage resources and activities supporting the event and to identify possible operational synergies; implementation of a new IT tool to facilitate the live integration of information
6. Healthcare	Coordination of operational programming activities in healthcare implemented by administrations and entities involved in the Jubilee, and integration with the planning made by the 7 other working groups	Technical Secretariat for the Jubilee; Municipality of Rome; Lazio Region; Croce Rossa Italiana; ARES—118; UNITALSI	Healthcare Plan, to manage first aid and emergency activities, as well as to monitor and manage infective diseases
7. Mobility	Shared planning of the interventions by transportation companies during Jubilee events; identification and monitoring of mobility needs, especially for people with disabilities	Technical Secretariat for the Jubilee; Municipality of Rome; Metropolitan City of Rome; Lazio Region; Local Police; Agency for Mobility; ANAS; ATAC; COTRAL; ENAC; Civitavecchia Port Authority; Trenitalia; Grandi Stazioni; Autostrade per l'Italia; Strade dei Parchi; RFI	Mobility Plan, to ensure an efficient management of the transportation services: the interventions concerned the Urban Network (improving the streets in the <i>Grande Viabilità Capitolina</i>), the rail transport (the railway transportation to and from San Pietro's train station was strengthened), the pedestrian mobility (new pedestrian paths were realized, and existing ones were improved), the management and control system for touristic buses and intelligent systems to control traffic
8. Italy—Vatican City relations	Supporting the activities carried out within the single working groups	n.a.	n.a.

intense effort to coordinate and manage the single events. Along with this, the coordinating activity also occurred in specific pre- and post-meetings, called debriefings, where the main results and issues of the events were analyzed and discussed, and in the Service Conferences. Given the small amount of ‘large Jubilee events’, debriefings and Service Conferences were few compared to the standard meetings of working groups.

As shown in Table 18.2, the Technical Secretariat for the Jubilee (TSG) set up all the tools needed to elaborate the best strategies to manage and coordinate the event. In particular, the TSG developed specific planning documents (outputs of the working groups): the Communication Plan, the Mobility Plan, the Safety Plan, the Healthcare Plan, and the Essential Services Plan.

We will now provide a graphical description of flows and relationship intensity (Social Network Analysis approach) for the ‘Mobility’ working group and for the ‘Communication and Infomobility’ subgroup (included in the ‘Media’ working group), in the perspective of evaluating the Technical Secretariat’s performance in managing and coordinating public networks (Fig. 18.3). For the sake of simplicity, Fig. 18.3 only takes into consideration the main actors, identified on the basis of their actual presence at meetings. The number in the arrow indicates the quantity of meetings occurred between the different stakeholders.

The figures confirm the centrality of the Prefecture of Rome (and of the Technical Secretariat for the Jubilee) as a strategic coordinator of the network of actors and as a key pillar of the governance model together with the Municipality of Rome, Lazio Region, and the Metropolitan City of Rome.

The high number of actors involved in the TSG (the number of actors invited to participate to the working groups), especially when evaluated in relation to the intensity of meetings and to the results, can be seen as a measure of the efficiency of the strategic coordination activity performed by the TSG. The continuous and small amount of meetings of the eight working groups (every 3 weeks or monthly), and the small amount of meetings between the TSG and the other actors (Fig. 18.3),

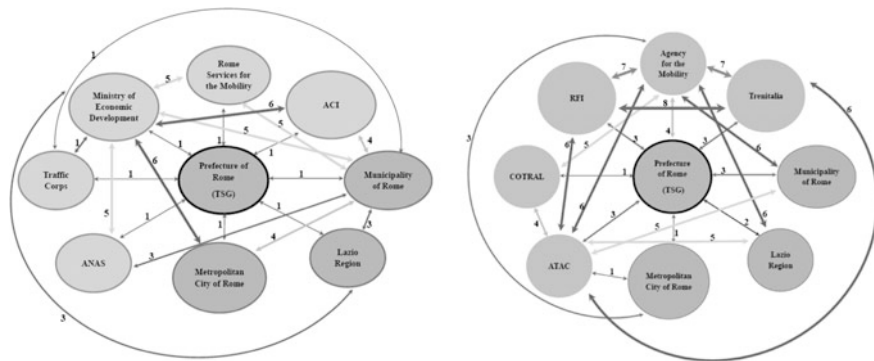


Fig. 18.3 Stakeholders’ meetings (‘Communication and Infomobility’ subgroup and ‘Mobility’ working group). *Source* own elaboration

show that the coordinating activity performed by the TSG was clear and effective. Despite the high complexity of the event (timeframe, different planning and management scenarios, amount of actors involved and engaged), the TSG was able to avoid the risks of a governance characterized by an excessive number of meetings and working groups (which would have often been inconclusive, partly because of rigid bureaucratic procedures to be followed).

The Technical Secretariat was also the operative tool through which the Prefect engaged public, private for-profit, and nonprofit stakeholders. The Jubilee, aside from being a partnership between two countries (Italy and the Vatican City State), also involved catholic nonprofit associations, social cooperatives, and foundations. The Jubilee was an extraordinary opportunity for 1000 volunteers from the national civil service, 2000 volunteers from the Rome Civil Protection, 'skilled volunteers' in the healthcare sector (500 volunteers) and, above all, 4000 catholic volunteers, directly coordinated by the Catholic Church.

The reason, as stated in Sect. 2, is that the governance activity requires the ability to design and manage formal and informal networks between the various public administrations, by combining models of collaborative networks and governance networks (Mandell 1999) (inter-institutional governance) and also to develop public-private partnerships (external governance) in order to mobilize financial resources outside the public sector.

It is important to highlight two priorities arising from the inter-institutional project management system implemented for the Jubilee, which represents a radical change from the previous experience in 2000.

The first priority is the importance of sharing knowledge between public administrations. European and Italian experiences in the field of local governance are an example of this: they had the goal of activating networks between public and private actors through the creation and exchange of knowledge, information, skills, and relationships ('network of networks' model and territorial creativity clusters).

The second priority is the choice to introduce a sort of 'integrated strategic agenda' for the network governance, aimed at supporting strengths obtained from the inter-institutional cooperation and from experimenting with public-private partnerships and inter-sector policies (healthcare, transportation, communication, volunteering, etc.)

In the planning phase of the Jubilee, the process was driven by a cartographic product based on sharing technical applications of various institutions (*Agenzia per la Mobilità, Metropolitan City of Rome, Protezione Civile Comune di Roma, Sistema Integrato Roma Sicura, Sogei, Lazio Region*), called IRIN system.

This allowed to create an integrated and dynamic planning, which could become a best practice and a potential model to be adopted for future 'complex' events in the Roman metropolitan context and at national (Italy) level. This dynamic approach to event management is a flexible and open tool by which single institutions and inter-institutional cooperation networks in the area bring their projects and policies together, continuously consolidating their integrated planning to employ the available resources, in a public value creation perspective.

IRIN was used by the inter-institutional ‘White Unique Room’ (*Sala Unica*), formed by all public administrations that managed safety for Jubilee events and requiring networking between ten Operative Rooms (*Sale Operative*) at the various institutions in the municipal area (*Questura di Roma, Comando Provinciale dei Carabinieri, Vigili del Fuoco, Guardia di Finanza, Corpo Forestale, Comando Polizia Municipale Roma Capitale, Protezione Civile Comune di Roma, ARES 118, ATAC, Agenzia per la Mobilità, Sovraintendenza Archeologica*).

The following initiatives contributed to strengthening this operational integration: a radio system to enforce integrated communication between the operative rooms of Local Police, the National Fire Corps and municipal utilities; the full integration of video-surveillance systems of all single administrations (5000 cameras); sharing all information platforms on urban (*Agenzia per la Mobilità*) and suburban (*ANAS and Autostrade per l’Italia*) mobility.

18.5 Preliminary Conclusions: Towards an Integrated Performance Evaluation System

To conclude, we can answer the research questions stated in Sect. 1 of this chapter.

For what concerns RQ1 (*Is it possible to adopt stakeholder mapping systems in the analysis of inter-institutional public networks?*), the analysis of the network of actors involved in the management of the event ‘Extraordinary Jubilee of Mercy’ allows to positively answer this question. Indeed, the perspectives of stakeholder management and multistakeholder engagement are important to understand the complexity of the system of the participating actors. The map of the different stakeholders who were engaged in the realization of the Jubilee of Mercy (Fig. 18.4), including both private and public agents, starts from the stakeholders’ classification based on the traditional criteria of power, legitimacy, and urgency (Freeman 1984; Phillips and Freeman 2010).

Other contributions (Bryson 2011; Fiorani et al. 2012; Freeman et al. 2010) could be adopted for clustering the stakeholders for the Extraordinary Jubilee of Mercy in four main categories: suppliers of goods and services, government institutions with a role of regulation, inter-institutional cooperation systems which deliver tangible and intangible services for the implementation of the event at zero-cost, civil society and nonprofit sector.

Concerning RQ2 (*Which public policy implementation approaches are the most appropriate for the management of highly complex events and inter-institutional projects?*), the case of the Extraordinary Jubilee of Mercy highlights the need to adopt a syncretic public policy implementation approach. Indeed, the syncretic approach adopted by the Technical Secretariat for the Jubilee achieved important results in terms of continuously rebuilding the network of interested actors and developing a collaborative network. Several issues were solved, such as those linked to the traditional institutional fragmentation at local level, the complexity of

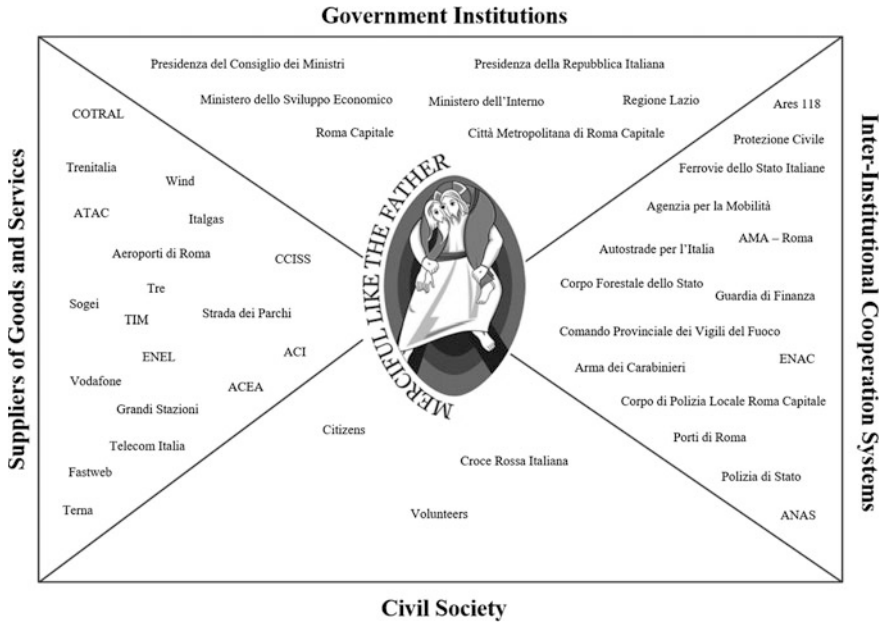


Fig. 18.4 Stakeholder Map for the Extraordinary Jubilee of Mercy. Source own elaboration

the network of actors involved, the strong cultural resistance of several stakeholders, both public and private, the implementation of systemic actions, and the lack of financial resources. Despite the high complexity of the event, the TSG was able to avoid a governance process that could be influenced by too many crowded working groups and meetings. Other methodologies, such as the implementation scenarios and the evaluability assessment, largely adopted in the US administration since the 80s (Poister 1981), could provide useful indications for the management of large inter-institutional events.

Regarding RQ3 (*Were project management tools effective in the case of the Extraordinary Jubilee of Mercy?*), several tools and methods were used, related to both project management (e.g., WBS, PERT, critical path method, critical chain, risk management) and impact evaluation, at social and economic level. In Fig. 18.5, we propose an EMBOK model for this case. An important factor to be mentioned is the limited timeframe: for example, the Jubilee of 2000 was planned and implemented (first two phases of the EMBOK model) over 6 years, while the 2016 Jubilee was announced only 8 months before the launch of the event (this also contrasted the process of finding funds). The case confirms that using project management strategies and theories for events brings the advantages highlighted by O’Toole and Mikolaitis (2002), mentioned in Sect. 2. The conceived and implemented project management systems have been influenced by a decision-making process that was flexible, reactive, inclusive, shared, and open to innovation, overcoming many of the limitations underlined by Bowdin et al. (2011).

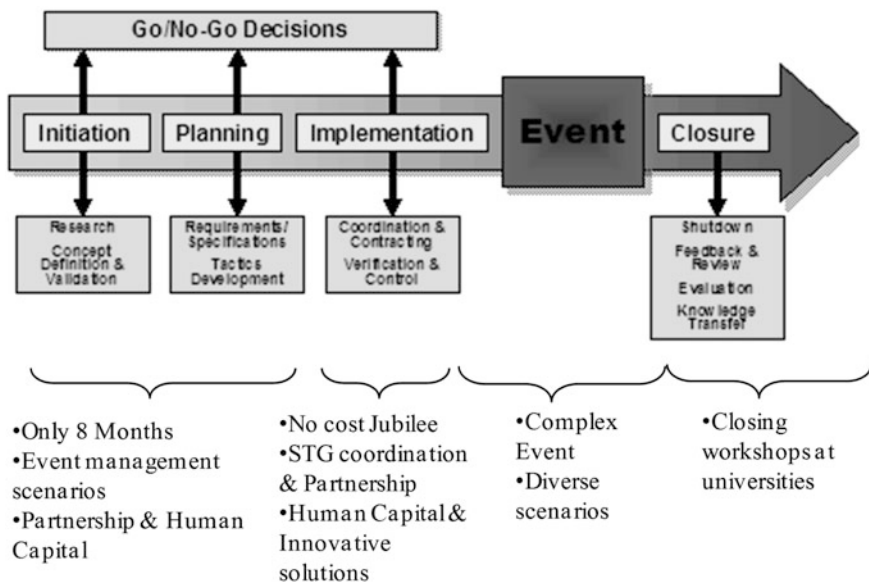


Fig. 18.5 EMBOK model for the Jubilee. Source own elaboration

In conclusion, the case of the Extraordinary Jubilee of Mercy, in our opinion, provides useful theoretical and empirical indications in the field of public management and governance.

The first concerns the definition of an integrated system for the evaluation of an outcome/trust-based inter-institutional project, which represents an important conceptual and operational challenge for the Italian public administration system.

This integrated system could provide important indications for the future management of large events (sports, culture, promotion of the economy, and the image of the country) and for the management of intervention policies facing environmental emergencies, such as recent earthquakes in Central Italy.

The Bouckaert and Halligan model represents a useful basis for the definition of this system of inter-institutional performance management (Bouckaert and Halligan 2008).

The Bouckaert and Halligan model distinguishes between width and depth of performance.

The experience of the Technical Secretariat lies between the meso level, which includes the management of the large event connected with the official public rational plans, and the macro level (from the central administration to the Region and the Municipality of Rome) with significant implications for international relations between the Italian Government and the Vatican City.

In this perspective, the horizontal dimension of width is very interesting, since it identifies different levels, from resources (people, logistics, finances, ICT, technology) to activities (such as working groups meetings, service conferences,

debriefing meetings, institutional meetings), to intermediate outputs (such as the various Plans), and most importantly to final outcomes and impacts, to the creation of shared value on the territory and community.

The impact, the creation of shared value, and trust are important dimensions in the evaluation of the Italian public administration system, which is characterized by issues of corruption, risk management, and transparency enhancement.

As a consequence, a broad and complex system to manage and evaluate performance was created (Table 18.3), which identifies several performance indicators that are consistent with the dual meso and macro perspective.

The second relevant conclusion, as Fiorani and Di Gerio (2016) show, is related to the introduction of a multidimensional assessment of the impact of the event. Short- and medium-long-term aspects should be taken into consideration, such as social and environmental impacts and legacy, in line with United Nations' Sustainable Development Goals (SDGs).

The following System Thinking map (Fig. 18.5) shows some preliminary results of the studies on these topics, which could be further analyzed in the future through the development of a System Dynamics model in line with the perspective of dynamic performance management.

As Fig. 18.6 shows, the partnerships (SDG n. 17, "Partnership for the goals"), i.e., the inter-institutional collaborations between public administrations, nonprofits institutions, and enterprises, through the eight thematic working groups coordinated by the Technical Secretariat, together with the presence of strong institutions (SDG n. 16, "Peace, Justice and Strong Institutions"), are the key success factors of this Jubilee.

Table 18.3 Performance management indicators. *Source* own elaboration

Activities	Intermediate outputs	Impacts and outcomes	Trust	Creation of shared public value
Working roundtables Service conferences Report writing Team management Project management Debriefing meetings	Implementation of the Jubilee Management Room Plans (Communications, Transportation, Safety, Healthcare) Cartographic system Event management scenarios	Widespread use of SIGIS Use of Radio Tetra Promotion of a risk management culture Websites and social media Strengthening of the mobility network Co-design between institutions	High consideration for the needs of people with disabilities Improved skills in managing healthcare emergencies	Media center Pilgrim Helpdesk

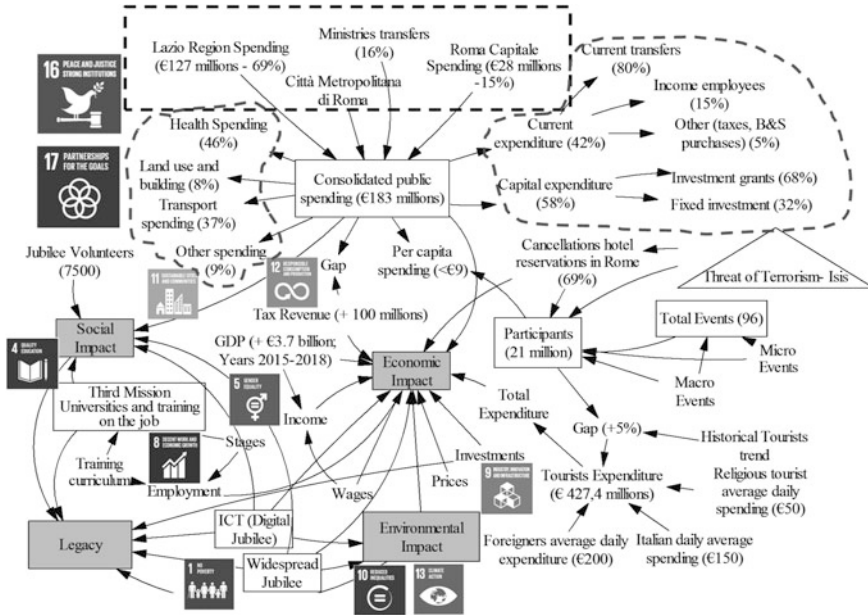


Fig. 18.6 Consolidated public spending and impacts: a sustainable Jubilee. Source own elaboration

It is worth to mention the collaboration with universities (SDG n. 4 “Quality Education”), which allowed to employ innovative managerial tools in planning, managing, and evaluating the impact of the event (social network analysis, system dynamics, and econometrics models).

In this regard, the University of Rome ‘Tor Vergata’, in close collaboration with the University of Rome ‘La Sapienza’, prepared two reports, the first on “consolidated public expenditures” (Dicorato et al. 2016) and the second on ‘Costs and Economic Impacts of the Jubilee’, aiming to evaluate direct and short-term economic impacts, as well as indirect or induced economic impacts (flow of visitors, increase in tourism and in consumption) (Ciccarone et al. 2016; Fiorani and Di Gerio 2016).

It is important to underline that the TSG, in cooperation with partner universities, organized several workshops on the wrap-up of the event, to promote the collection of feedback and ideas on the legacy of the Jubilee. These workshops fostered discussions between different actors, directly and indirectly involved in the event, as well as the transfer of accumulated knowledge.

Finally, we should note that TSG’s actions were driven by the principles of saving and cost-effectiveness of spending (SDG n. 12, “Responsible Production and Consumption”), thanks to sharing human resources and available tools. This process prevalently focused on private investments (sponsorships), and also on ordinary resources of each entity (Fig. 18.6 shows that the consolidate public spending was 183 million €), which led to the definition of a “zero-cost Jubilee”.

References

- Agranoff, R., & McGuire, M. (2001). Big questions in public network management research. *Journal of Public Administration Research and Theory*, 11(3), 295–396.
- Bardach, E. (1977). *The implementation game. What happens after a bill becomes a law*. Cambridge: MIT.
- Bouckaert, G. (2014). Coordinating hierarchies, markets, and networks: a catch 22? TAD-Transatlantic Dialogue, Lugano, Switzerland, 5–7 June.
- Bouckaert, G., & Halligan, J. (2008). *Managing performance: International comparison*. London: Routledge.
- Bovaird, T., & Löffler, E. (2003). Evaluating the quality of public governance: Indicators models and methodologies. *International Review of Administrative Sciences*, 69(3), 313–328.
- Bowdin, G., Allen, J., O'Toole, W. G., Harris, R., & McDonnell, I. (2006). *Events management* (2nd ed.). Oxford: Butterworth-Heinemann.
- Bowdin, G., Allen, J., O'Toole, W. G., Harris, R., & McDonnell, I. (2011). *Events management* (3rd ed.). Oxford: Butterworth-Heinemann.
- Bryson, J. (2011). *Strategic planning for public and nonprofit organizations* (4th ed.). San Francisco: Jossey-Bass.
- Cepiku, D. (2005). Governance: riferimento concettuale o ambiguità terminologica nei processi di innovazione della PA? *Azienda Pubblica*, 18(1), 105–131.
- Ciccarone, G., Di Bartolomeo, G., Di Simone, V., Fedeli, S., & Tancioni, M. (2016). L'impatto macroeconomico del Giubileo della Misericordia sull'area romana. In A. Tramontano (Ed.), *Giubileo Straordinario della Misericordia: L'esperienza della Segreteria Tecnica per il Giubileo*. Istituzioni, Università e territorio nell'Anno Santo della Misericordia.
- Clarke, M., & Stewart, J. (1997). *Handling the wicked issues: A challenge for government*. Birmingham: Institute of Local Government.
- Dente, B. (1982). L'analisi dell'efficacia delle politiche pubbliche: problemi di teoria e di metodo. *Rivista Trimestrale di Scienza dell'Amministrazione*, 4, 105–133.
- Dicorato, S. L., Marchese, B., Paciullo, G., & Meneguzzo, M. (2016). Il consolidamento della spesa delle amministrazioni pubbliche per la realizzazione del Giubileo della Misericordia. In A. Tramontano (Ed.), *Giubileo Straordinario della Misericordia: L'esperienza della Segreteria Tecnica per il Giubileo, Segreteria Tecnica per il Giubileo*. Istituzioni, Università e territorio nell'Anno Santo della Misericordia.
- Eljassen, K. A., & Kooiman, J. (Eds.). (1993). *Managing public organisations: lessons from contemporary European experience*. London: Sage.
- Ferdinand, N., & Kitchin, P. (2012). *Events management: An international approach*. London: Sage.
- Ferlie, E., Lynn, L. E., & Pollitt, C. (Eds.). (2005). *The Oxford handbook of public management*. Oxford: Oxford University Press.
- Fiorani, G., & Di Gerio, C. (2016). Valutazione di impatto multidimensionale del Giubileo della Misericordia. In A. Tramontano (Ed.), *Giubileo Straordinario della Misericordia: L'esperienza della Segreteria Tecnica per il Giubileo, Segreteria Tecnica per il Giubileo*. Istituzioni, Università e territorio nell'Anno Santo della Misericordia.
- Fiorani, G., Meneguzzo, M., & Jannelli, R. (Eds.). (2012). *CSR 2.0 proattiva e sostenibile. Tra mercati globali e gestione della crisi*. Milan: Egea.
- Freeman, R. E. (1984). *Strategic management: A stakeholder approach*. Boston: Pitman.
- Freeman, R. E., Harrison, J. S., Wicks, A. C., Parmar, B. L., & De Colle, S. (2010). *Stakeholder theory: The state of the art*. Cambridge: Cambridge University Press.
- Getz, D. (1991). *Festivals, special events, and tourism*. New York: Van Nostrand Reinhold.
- Getz, D. (2005). *Event management and event tourism* (2nd ed.). New York: Cognizant.

- Hall, M. (1992). *Hallmark tourist events: Impacts, management and planning*. London: Belhaven.
- Hanf, K., & Scharpf, F. W. (1978). *Interorganizational policy making*. London: Sage.
- Isett, K. R., Mergel, I. A., LeRoux, K., Mischen, P. A., & Rethemeyer, R. K. (2011). Networks in public administration scholarship: Understanding where we are and where we need to go. *Journal of Public Administration Research and Theory*, 21(Suppl 1), i157–i173.
- Jackson, N. (2013). *Promoting and marketing events: Theory and practice*. New York: Routledge.
- Jago, L. K., & Shaw, R. N. (1998). Special events: A conceptual and differential framework. *Festival Management and Event Tourism*, 5(1/2), 21–32.
- Kickert, W. J. M., Klijn, E. H., & Koppenjan, J. F. M. (1997). *Managing complex networks*. London: Sage.
- Klijn, E. H. (2010). Trust in governance networks: Looking for condition for innovative solutions and outcome. In S. P. Osborne (Ed.), *The New Public Governance? Emerging perspectives on the theory and practice of public governance*. London: Routledge.
- Kuhlmann, S. K., & Wayenberg, E. (2016). Institutional impact assessment in multi-level systems: Conceptualizing decentralization effects from a comparative perspective. *International Review of Administrative Sciences*, 82(2), 233–254.
- Mandell, M. P. (1999). The impact of collaborative efforts: Changing the face of public policy through networks and network structures. *Policy Studies Review*, 16(1), 4–17.
- Meneguzzo, M. (1995). Dal New Public Management alla Public Governance: il pendolo della ricerca sulla amministrazione pubblica. *Azienda Pubblica*, 3(1), 5–20.
- Meneguzzo, M., & Del Vecchio, M. (Eds.). (2000). *Metodi di analisi e valutazione dei programmi pubblici*. Milano: Egea.
- Milward, H. B., & Provan, K. G. (2003). Managing the Hollow-State: Collaboration and contracting. *Public Management Review*, 5(1), 1–18.
- Monnier, E. (1999). Methods for partnership evaluation. In *Speech for the European conference in Berlin*, MEANS.
- Mussari, R. (1999). *La valutazione dei programmi nelle aziende pubbliche*. Torino: Giappichelli.
- Nakamura, R. T., & Smallwood, F. (1980). *The politics of policy implementation*. New York: St. Martin's Press.
- O'Toole, W. J. (2000). Towards the integration of event management best practices by the project management process. In J. Allen, R. Harris, L. K. Jago, & A. Veal (Eds.), *Events beyond 2000: Setting the agenda*. Sydney: Australian Centre for Event Management.
- O'Toole, W. J., & Mikolaitis, P. (2002). *Corporate event project management*. New York: Wiley.
- Osborne, S. P. (Ed.). (2010). *The New Public Governance? Emerging perspectives on the theory and practice of public governance*. London: Routledge.
- Phillips, R. A., & Freeman, R. E. (2010). *Stakeholders*. Camberley: Edward Elgar.
- Poister, T. H. (1981). Program evaluation: State of the art/ state of the practice. *Public Administration Review*, 41(2), 280.
- Pollitt, C., & Bouckaert, G. (2011). *Public management reform: A comparative analysis* (3rd ed.). Oxford: Oxford University Press.
- Pollitt, C., & Hupe, P. (2011). Talking about government: The role of magic concepts. *Public Management Review*, 89(1), 114–127.
- Pressman, J. L., & Wildavsky, A. (1973). *Implementation*. Berkeley: University of California Press.
- Provan, K. G., & Kenis, P. (2008). Modes of network governance: Structure, management, and effectiveness. *Journal of Public Administration Research and Theory*, 18(2), 229–252.
- Raj, R., & Musgrave, J. (2009). *Event Management and Sustainability*. London: CAB International.
- Rebora, G. (1989). L'analisi strategica come strumento di management pubblico. *Economia & Management*.

- Shone, A., & Parry, B. (2004). *Successful event Management: A practical handbook*. London: Thomson Learning.
- Silvers, J. R. (2008). *Risk management for meetings and events*. Oxford: Butterworth-Heinemann.
- Silvers, J. R., Bowdin, G., O'Toole, W. J., & Nelson, K. B. (2006). Towards an international event management body of knowledge (EMBOK). *Event Management*, 9(4), 185–198.
- Thomas, O., Hermes, B., & Loos, P. (2008). Reference model-based event management. *International Journal of Event Management Research*, 4(1), 38–57.
- Wildavsky, A. (1981). Per un'analisi delle public policies. *Queste Istituzioni*.

Part V
Dealing with the Challenges of
Outcome-Based Performance
Management: Innovative Methods and
Tools

Chapter 19

Performance Benchmarking of School Districts in New York State

Thomas R. Sexton, Christie Comunale, Michael Shane Higuera
and Kelly Stickle

Abstract We used DEA to measure the performance of New York State school districts and provide alternative improvement targets for each district. We found that 201 of the 624 (32.2%) districts with one or more high schools and 28 of the 31 (90.3%) districts with no high school were on the performance frontier. We found evidence that NYS could reduce FTE teachers by 8.4%, FTE teacher support by 17.2%, and FTE administration and professional staff by 9.4%. We also found that NYS could increase percentage of students who pass the English exam by 4.9 percentage points, the mathematics exam by 5.0 percentage points, and the science exam by 5.8 percentage points, while increasing the average graduation rate by 5.4 percentage points.

Keywords School districts · Benchmarking · Data envelopment analysis · New York State

19.1 Introduction

In 2011, New York State's 695 school districts (New York State Education Department n.d.) spent \$53.7 billion (U.S. Census Bureau 2011, Table 6) to educate almost 2.7 million elementary and secondary pupils (U.S. Census Bureau, 2011, Table 19), a cost of over \$19,000 per pupil (U.S. Census Bureau 2011, Table 8). Elementary and secondary education accounts for nearly one-quarter of all state and local expenditures in New York State (U.S. Government Spending n.d.). While New York State has some excellent school districts, others struggle with poor standardized test scores and low graduation rates. Many of the reasons for the differences among school districts are widely accepted. These include differences in wealth, English proficiency, and inefficient use of resources.

T.R. Sexton (✉) · C. Comunale · M.S. Higuera · K. Stickle
College of Business, Stony Brook University, 317 Harriman Hall,
Stony Brook, NY 11794-3775, USA
e-mail: Thomas.Sexton@stonybrook.edu

Given the high cost of public education and its critical importance for the future of New York and the nation, it is natural for taxpayers, legislators, and administration officials to hold public education institutions accountable for producing high quality outcomes. To do so, we must measure the performance of each school district in an objective, data-informed manner. Commonly used methods for performance measurement under these circumstances are often called *benchmarking models*. When applied to school districts, a benchmark model identifies leading school districts, called *benchmark school districts*, and it facilitates the comparison of other school districts to the benchmark school districts. Nonbenchmark school districts can focus on specific ways to improve their performance and thereby that of the overall statewide school system.

In this paper, we utilize Data Envelopment Analysis (DEA) to measure the performance of New York State school districts in the 2011–2012 academic year, and provide detailed alternative improvement pathways for each school district.

19.2 Literature Review

DEA has been used since the 1950s in a wide variety of applications, including health care, banking, pupil transportation, and most recently, education. DEA's mathematical development may be traced to Charnes et al. (1978), who built on the work of Farrell (1957) and others. The technique is well documented in the management science literature (Charnes et al. 1978, 1979, 1981; Sexton 1986; Sexton et al. 1986; Cooper et al. 1999), and it has received increasing attention as researchers have wrestled with problems of productivity measurement in the services and nonmarket sectors of the economy. Emrouznejad et al. (2008) provided a review of more than 4000 DEA articles. See Emrouznejad (2014) for an extensive bibliography of DEA publications as well as a DEA tutorial and DEA software.

We are not the first to apply DEA to school districts. Färe et al. (1989) applied DEA to evaluate the performance of a sample of Missouri school districts for the 1985–1986 school year. Kirjavainen and Loikkanen (1998) studied efficiency differences among Finnish senior secondary schools. They found that schools with small classes and heterogeneous student bodies were inefficient whereas school size did not affect efficiency. Surprisingly, private schools were inefficient relative to public schools. Kang and Greene (2002) used DEA to evaluate the impacts of institutional arrangements on various measures of high school output. Driscoll et al. (2003) used statistical methods to estimate a production function for California schools, which showed that smaller districts, smaller schools, and smaller class sizes were associated with higher academic achievement scores. Ruggiero (2007) used DEA to examine efficiency, costs and adequacy of 607 Ohio school districts using school year 2000 data. The results indicate that adequacy standards can be met by improving the performance of inefficient school districts and reallocating existing resources without increasing total expenditures. Thanassoulis et al. (2016)

review applications of DEA in secondary and tertiary education, focusing on the opportunities that this offers for benchmarking at institutional level.

19.3 Data Envelopment Analysis

DEA has proven to be a successful tool in performance benchmarking. It is particularly well suited when measuring the performance of units along multiple dimensions, as is the case with complex organizations such as school districts. DEA empirically identifies the best performers by forming the performance frontier based on observed indicators from all units. Consequently, DEA bases the resulting performance scores and potential performance improvements entirely on the actual performance of other DMUs, free of any questionable assumptions regarding the mathematical form of the underlying production function. On balance, many analysts view DEA as preferable to other forms of performance measurement.

Figures 19.1 and 19.2 illustrate the performance frontier for a simple model of school districts. We can use this simple model, which is clearly inadequate for capturing the complexity of school districts, to demonstrate the fundamental concepts of DEA. In this model, we assume that each school district employs only one type of resource, full-time equivalent (FTE) teachers, and prepares students for only one type of standardized test, mathematics at the appropriate grade level, measured

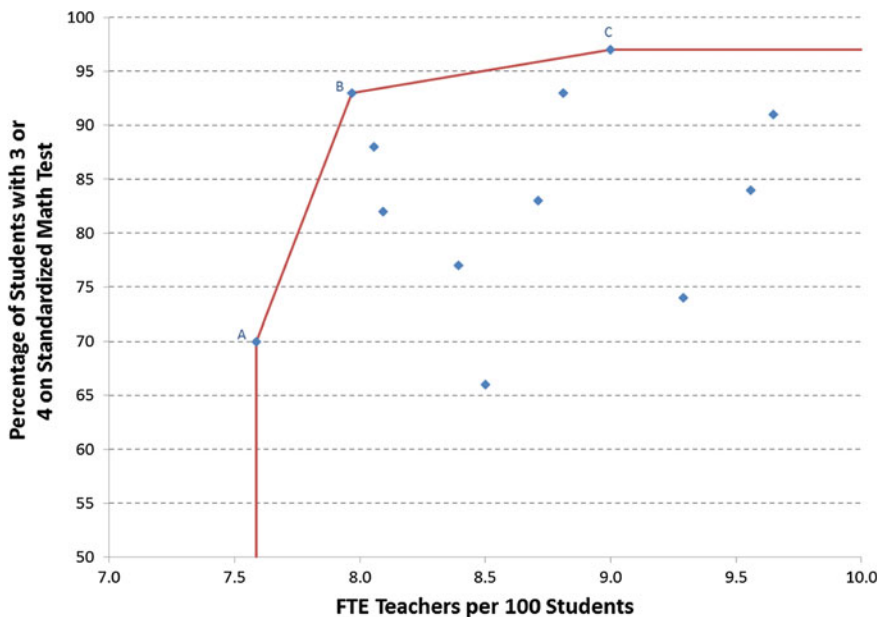


Fig. 19.1 The performance frontier for a simple example

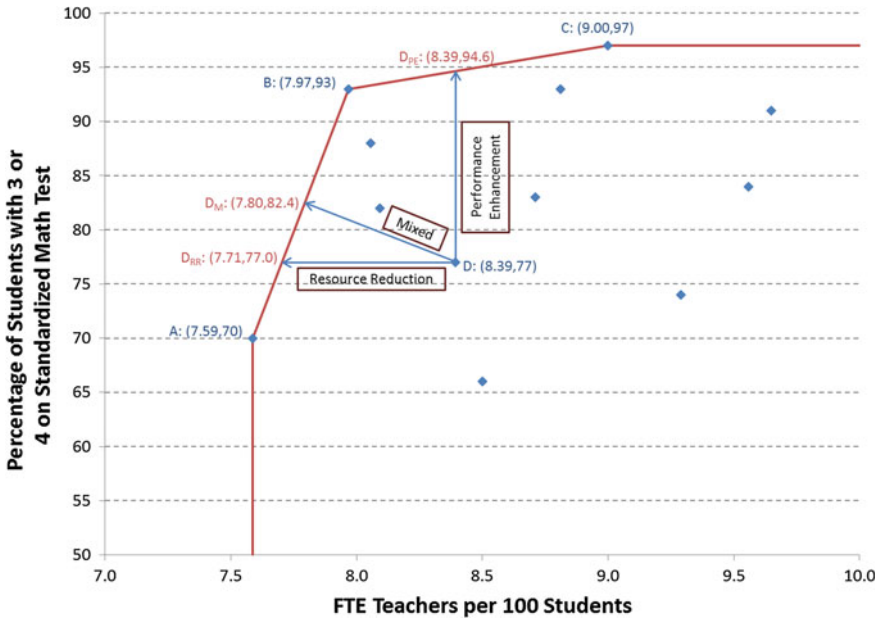


Fig. 19.2 Several ways for school district D to move to the performance frontier

as the percentage of students who score at a given level or higher. Each school district is represented by a point in the scatterplot.

In Fig. 19.1, school districts A, B, and C define the performance frontier. In each case, there is no school district or weighted average of school districts that has fewer FTE teachers per 100 students and has a higher percentage of students who scored 3 or 4 on the standardized mathematics test. Such school districts, if they existed, would lie to the Northwest of A, B, or C, and no such districts, or straight lines between any two districts, exists.

School district D, in Fig. 19.2, does not lie on the performance frontier and therefore its performance can improve. In principle, D can choose to move anywhere on the performance frontier. If school district D chooses to focus on resource reduction without test performance change, it would move to the left, reaching the performance frontier at point D_{RR}. This move would require a reduction from 8.39 to 7.71 FTE teachers per 100 students. If school district D enrolls 10,000 students, this reduction would be from 839 to 771 teachers, a percentage reduction of 8.1%. We refer to this strategy as the *resource reduction orientation*.

If school district D chooses to focus on performance enhancement without resource reduction, it would move upward, reaching the performance frontier at point D_{PE}. This move would require 94.6% of its students to score 3 or 4 on the standardized mathematics test, up from 77%. If 1000 students in school district D sat for the standardized mathematics test, students scoring 3 or 4 would increase would from 770 to 946, or by 22.9%. We refer to this strategy as the *performance enhancement orientation*.

School district D might prefer an intermediate approach that includes both resource reduction and performance enhancement and move to point D_M . This entails both a reduction in FTE teachers per 100 students from 8.39 to 7.80 and an increase in the percentage of students who score 3 or 4 on the standardized mathematics test from 77 to 82.4%. If school district D enrolls 10,000 students, this reduction would be from 839 to 780 teachers, or by 7.0%, and an increase in students scoring 3 or 4 from 770 to 824, or 7.0%. We refer to this strategy as the *mixed orientation*. The mixed orientation has the feature that the percentage decrease in each resource equals the percentage increase in each performance measure.

The three points D_{RR} , D_{PE} , and D_M are called *targets* for school district D because they represent three possible goals for D to achieve to reach the performance frontier. School district D can choose its target anywhere on the performance frontier, but these three points represent reasonable reference points for D as it improves its overall performance.

Of course, this model does not consider other resources used by school districts such as teacher support personnel and other staff, nor does it consider standardized test scores in science or English. It also ignores graduation rates in school districts with one or more high schools. Moreover, it does not recognize differences in important district characteristics such as the number of elementary and secondary students, the percentage of students who qualify for free or reduced price lunch or who have limited English proficiency, or the district's combined wealth ratio.

When other measures are included in the model, we can no longer rely on a simple graphical method to identify a school district's target school district. For this purpose, we rely on the linear programming model that we describe in detail in Technical Appendix. Nonetheless, the target school district will have the same basic interpretation. Relative to the school district in question, the target school district consumes the same or less of each resource, its students perform the same or better on each standardized test, its graduation rate is at least as high (if applicable), it educates the same number or more students, and it operates under the same or worse district characteristics.

19.4 A DEA Model for School District Performance in New York State

To apply the DEA methodology to measure the performance of New York State school districts, we began by identifying three categories of important school district measurements. They were:

- resources consumed;
- performance measures; and
- district characteristics

We defined the resources consumed as:

- FTE teachers;
- FTE teacher support (teacher assistants + teacher aides); and
- building administration and professional staff (principals + assistant principals + other professional staff + paraprofessionals).

For school districts with no high school, we defined the performance measures as:

- percentage of students scoring at or above level 3 on ELA grade 6;
- percentage of students scoring at or above level 3 on math grade 6; and
- percentage of students scoring at or above level 3 on science grade 4.

For school districts with one or more high schools, we defined the performance measures as:

- total cohort results in secondary level English after 4 years of instruction: percentage scoring at levels 3–4;
- total cohort results in secondary level math after 4 years of instruction: percentage scoring at levels 3–4;
- grade 8 science: percentage scoring at levels 3–4 all students; and
- 4-year graduation rate as of August.

We defined the district characteristics as:

- number of elementary school students;
- number of secondary school students;
- percentage of students with free or reduced price lunch;
- percentage of students with limited English proficiency; and
- school district's combined wealth ratio.

We recognize that other choices of variables are possible. We use this particular set of variables because it captures a reasonable range of resources consumed, performance dimensions to be measured, and district characteristics to be taken into account. Other variables may be added if statewide data are available for every school district. Our objective is to illustrate the model and its ability to provide school districts with useful feedback for strategic planning and other purposes.

Our goal is to guide school district managers to the performance frontier and we recognize that there are infinitely many points on the performance frontier. We also believe that school district managers are in the best position to decide which direction to take toward the performance frontier, and that different school districts may choose different strategies based on their own circumstances.

Toward that end, we consider three possible orientations, or directions to the performance frontier, for each school district. The resource reduction orientation seeks to reduce resource consumption as much as possible while maintaining performance measures at their current levels. The performance enhancement orientation seeks to improve performance measures as much as possible while

maintaining resource consumption at current levels. The mixed orientation seeks to improve performance measures and reduce resource consumption simultaneously in a balanced way.

We present the results of all three orientations to provide school district administrators with alternative options for reaching the performance frontier. One district might elect to focus on resource reduction; another might opt for increases in test scores and graduation rate, while a third might prefer a blended strategy that combines these two objectives. Since there are infinitely many points on the performance frontier toward which a district may move, the three that we present are designed to highlight three possible alternatives.

We point out that the performance frontier is unaffected by the choice of orientation. Any district that lies on the performance frontier in one orientation will also lie on it in any other orientation. Orientation only determines the location of the target district on the performance frontier.

19.5 Data and Results

We obtained complete data for 624 public school districts with one or more high schools and 31 public school districts with no high school for the academic year 2011–2012. Complete data were unavailable for certain districts. All data were obtained from the New York State Education Department.

19.6 Results for Three Example Districts

Table 19.1 shows the results for three districts based on the model described above. These districts were selected to illustrate the manner in which the model results can be presented to school districts and how they might be interpreted.

School district A would reduce all three resources by 18.3% using the resource reduction orientation and by 4.0% under the mixed orientation, but would not reduce any resources under the performance enhancement orientation. Improvements in English and science would be virtually the same using all three orientations (in the range of 4%) but the improvements in math and graduation rate are notably higher using either the performance enhancement or mixed orientations. The message for school district A is that it can raise all three test measures by about 4% and graduation rate by about 8% with little or no reduction in resources. Alternatively, it can improve English and science (but not math) by about 4% and graduation rate by 4–5% even with significant resource reductions. The choice of strategy would be influenced by many other factors not reflected in the model.

School district B can reduce its FTE teachers by at least 6.9% but its greater opportunity lies in teacher support, which it can reduce by at least 27.4%. Despite these reductions, it can improve English by almost 7% and math by almost 4%.

Table 19.1 Results for three example districts under three orientations (in percentages of current actual values)

Dist	Orientation	FTE teachers	FTE teacher support	Bld adm and prof staff	Secondary level English (%)	Secondary level math (%)	Grade 8 science (%)	Grad rate (%)
A	Res red	81.7	81.7	81.7	103.9	100.0	103.8	104.6
	Perf enhan	100.0	100.0	100.0	104.3	104.3	104.3	108.5
	Mixed	96.0	96.0	96.0	104.3	104.0	104.0	108.2
B	Res red	90.2	65.8	90.2	105.3	101.5	100.0	100.0
	Perf enhan	93.1	72.6	100.0	106.8	103.8	101.8	101.8
	Mixed	92.8	72.6	98.4	106.7	103.6	101.6	101.6
C	Res red	99.7	99.7	99.7	101.1	113.8	100.0	100.9
	Perf enhan	100.0	100.0	100.0	101.1	113.8	100.1	101.1
	Mixed	99.9	99.9	99.9	101.1	113.8	100.1	101.0

The values in bold draw attention to specific areas for improvement

School district C is performing very well regardless of orientation with the exception of math, which it can improve by almost 14%.

19.7 Statewide Results

We found no evidence that 201 of the 624 (32.2%) school districts with one or more high schools can reduce resource consumption or improve performance. The same statement applies to 28 of the 31 (90.3%) school districts with no high school. Put another way, each of these school districts serves as its own target school district. Based on the observed performance of all school districts in New York State in the academic year 2011–2012, none of these school districts can simultaneously reduce *each* of its resources and improve *each* of its performance measures while operating under the same district characteristics.

It is important to recognize that DEA is an empirical method that measures a school district's performance *relative to the performances of other school districts*. It makes no theoretical assumptions about the shape of the production possibility frontier. Thus, it may in fact be possible for school districts on the frontier to make improvements but current data provide no evidence to suggest this possibility or indicate the extent of such improvements.

19.8 Districts With One or More High Schools

The 624 school districts with one or more high schools employed 126,470 FTE teachers, 33,035 FTE teacher support personnel, and 25,492.5 FTE building administration and professional staff in the academic year 2011–2012. The average

Table 19.2 Data and statewide results for all three orientations for school districts with one or more high schools

	FTE teachers	FTE teacher support	Building admin and prof staff	Secondary level English (%)	Secondary level math (%)	Grade 8 science (%)	Grad rate (%)
Actual	126,470	33,035	25,493	84.4	86.0	81.6	84.2
Mixed orientation							
Target	115,812	27,359	23,091	89.3	91.0	87.4	89.6
Change	10,658	5676	2402	4.9	5.0	5.8	5.4
% Change	8.4	17.2	9.4	5.8	5.8	7.1	6.4
Resource reduction orientation							
Target	<i>102,314</i>	<i>25,653</i>	<i>20,567</i>	86.7	88.4	85.3	86.5
Change	<i>24,156</i>	<i>7382</i>	<i>4925</i>	2.2	2.4	3.7	2.3
% Change	<i>19.1</i>	<i>22.3</i>	<i>19.3</i>	2.6	2.8	4.5	2.7
Performance enhancement orientation							
Target	119,311	27,913	23,687	89.7	<i>91.3</i>	<i>87.6</i>	<i>89.9</i>
Change	7159	5122	1805	5.3	<i>5.3</i>	<i>6.0</i>	<i>5.7</i>
% Change	5.7	15.5	7.1	6.3	<i>6.1</i>	<i>7.3</i>	<i>6.8</i>

The italicized values represent values in the direction of the model’s orientation

percentage of students who scored 3 or 4 on the English exam was 84.4%; on the mathematics exam, the average was 86.0%, and on the science exam, the average was 81.6%. The average graduation rate was 84.2%. See Table 19.2.

Using a mixed orientation, we found evidence that the number of FTE teachers can be reduced by 8.4%, the number of FTE teacher support personnel can be reduced by 17.2%, and the number of FTE building administration and professional staff personnel can be reduced by 9.4%. In addition, that the average¹ percentage of students who score 3 or 4 on the English exam can rise by 4.9 percentage points, by 5.0 percentage points on the mathematics exam, and by 5.8 percentage points on the science exam. Moreover, the average² graduation rate can rise by 5.4 percentage points.

Using a resource reduction orientation, we found evidence that the number of FTE teachers can be reduced by 19.1%, the number of FTE teacher support personnel can be reduced by 22.3%, and the number of FTE building administration and professional staff personnel can be reduced by 19.3%. In addition, the average percentage of students who score 3 or 4 on the English exam can rise by 2.2 percentage points, by 2.4 percentage points on the mathematics exam, and by 3.7 percentage points on the science exam. Moreover, the average graduation rate can

¹These are unweighted averages and therefore they do not represent the statewide percentages.

²See previous footnote.

rise by 2.3 percentage points. We point out that, even though we have used a resource reduction orientation, we can still identify potential improvements in the performance measures. This is because one or more of the performance measure constraints may not be binding at optimality; the corresponding slacks represent potential improvements known as nonradial improvements.

Finally, using a performance enhancement orientation, we found evidence that the number of FTE teachers can be reduced by 5.7%, the number of FTE teacher support personnel by 15.5%, and the number of FTE building administration and professional staff personnel by 7.1%. In addition, the average percentage of students who score 3 or 4 on the English exam can rise by 5.3 percentage points, by 5.3 percentage points on the mathematics exam, and by 6.0 percentage points on the science exam. Moreover, the average graduation rate can rise by 6.8 percentage points. Once again, even though we have used a performance enhancement orientation, we can still identify potential reductions in the resource measures. This is because one or more of the resource constraints may not be binding at optimality; the corresponding slacks represent potential nonradial reductions in resource consumption.

Figures 19.3, 19.4, and 19.5 illustrate the potential improvements in the three resource categories. For districts that lie on the diagonal of one of these graphs, there is no evidence that they could reduce their use of this resource category. Other districts have the potential to reduce resource consumption by the amount that they lay below the diagonal.

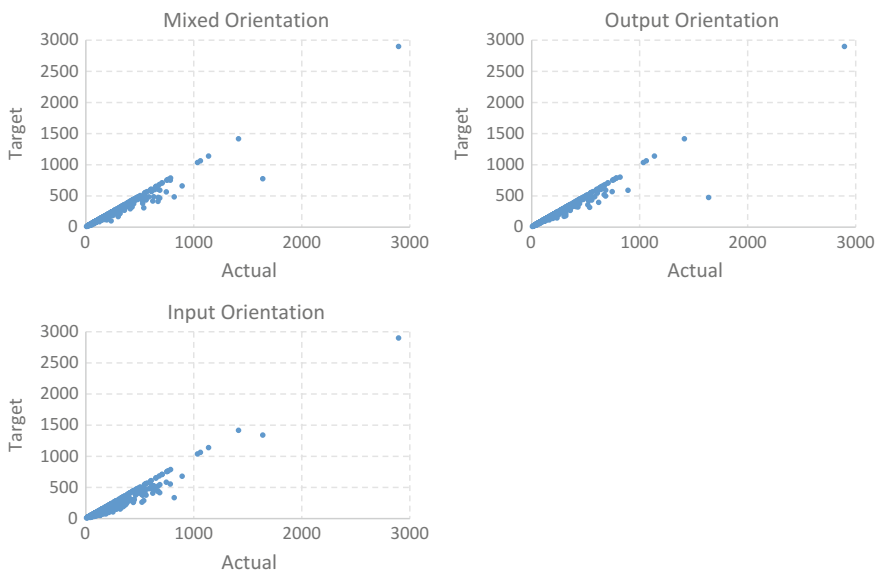


Fig. 19.3 Target versus actual FTE teachers under each of the three orientations for school districts with at least one high school

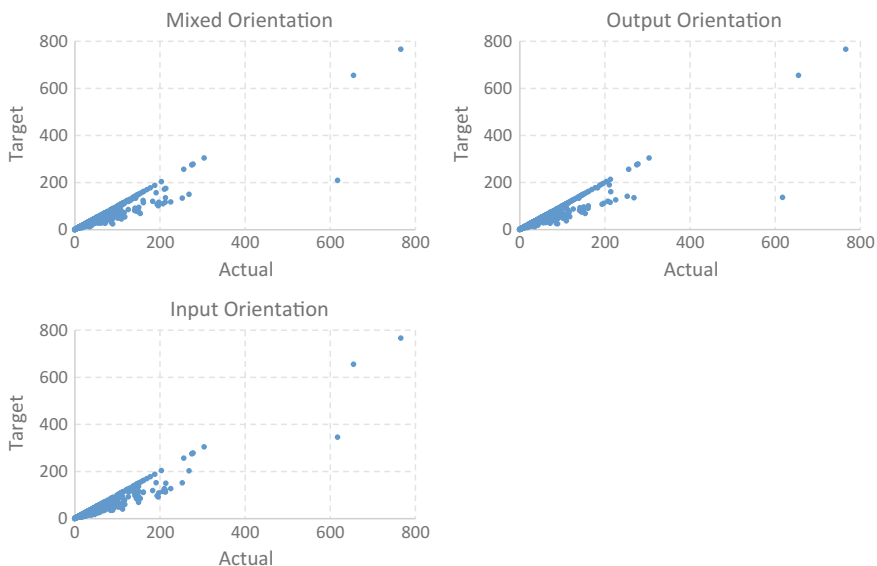


Fig. 19.4 Target versus actual FTE teacher support under each of the three orientations for school districts with at least one high school

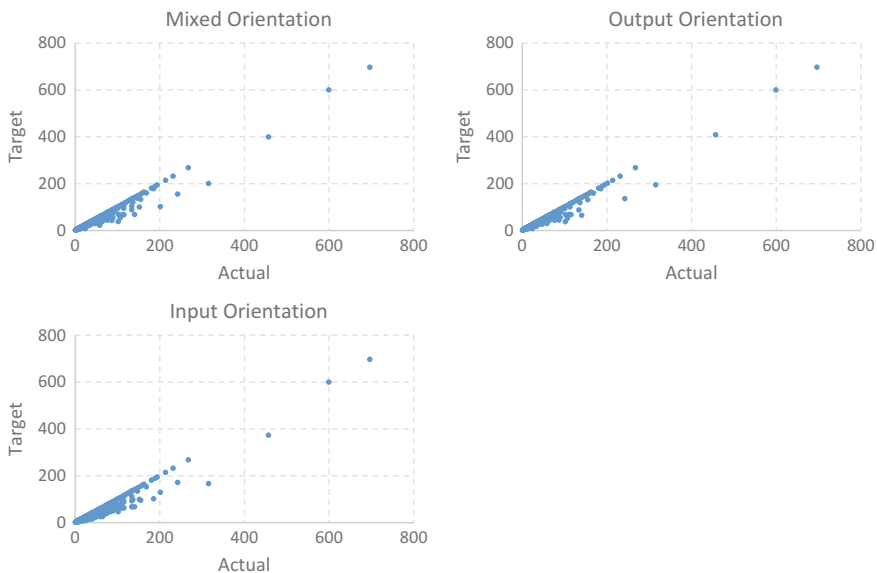


Fig. 19.5 Target versus actual FTE building and administrative professional staff under each of the three orientations for school districts with at least one high school

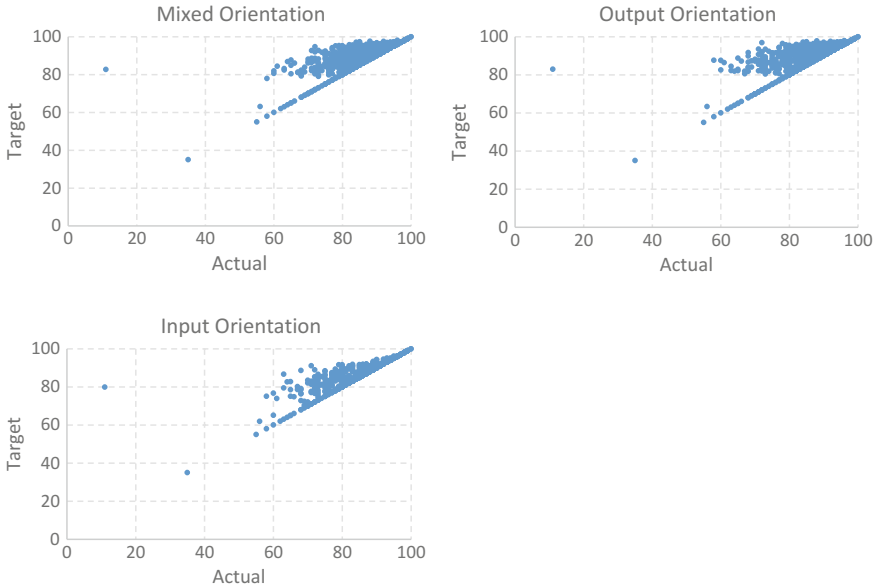


Fig. 19.6 Target versus actual percentage of students scoring 3 or 4 on the secondary level English standardized test under each of the three orientations for school districts with at least one high school

Figures 19.6, 19.7, 19.8 and 19.9 illustrate the potential improvements in the four performance measures. For districts that lie on the diagonal of one of these graphs, there is no evidence that they could improve their performance in this dimension. Other districts have the potential to improve by the amount that they lay above the diagonal.

Figure 19.10 shows the histograms of the school districts for each of the three factor performances associated with the resources, excluding those districts for which no improvement is possible. Figure 19.11 shows the histograms of the school districts for each of the four factor performances associated with the performance measures, again excluding those for which no improvement is possible.

19.9 Districts Without a High School

The 31 school districts with no high school employed 2233 FTE teachers, 762 FTE teacher support personnel, and 416 FTE building administration and professional staff in the academic year 2011–2012. The average percentage of students who scored 3 or 4 on the English exam was 84.4%; on the mathematics exam, the average was 86.0%, and on the science exam, the average was 81.6%. See Table 19.3.

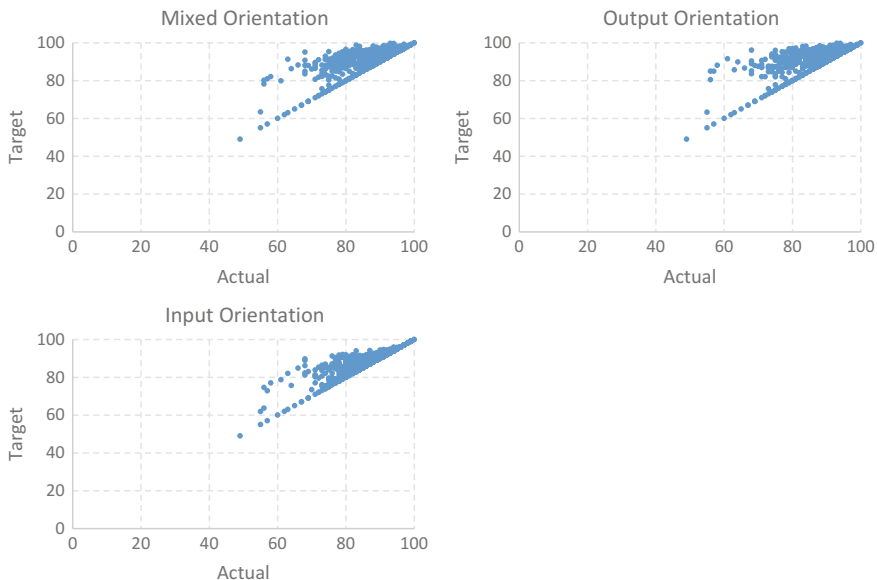


Fig. 19.7 Target versus actual percentage of students scoring 3 or 4 on the secondary level mathematics standardized test under each of the three orientations for school districts with at least one high school

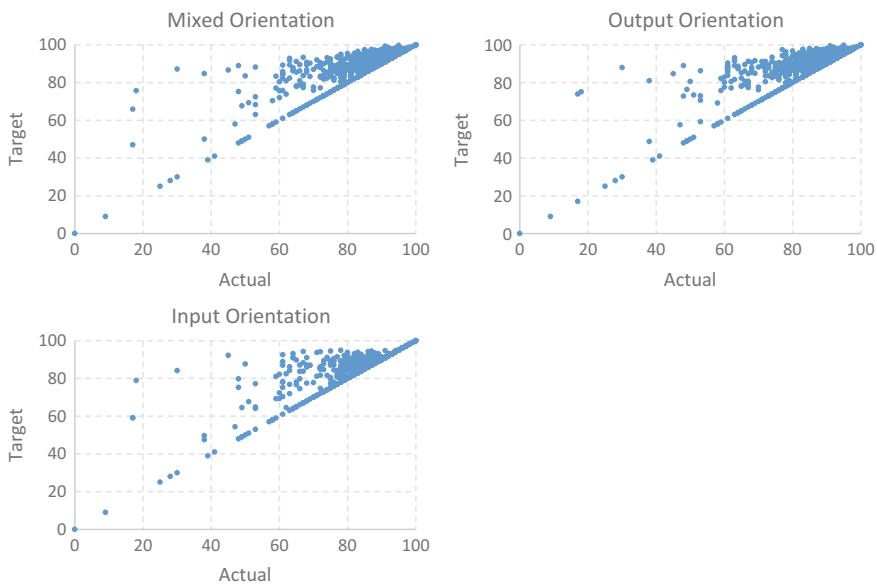


Fig. 19.8 Target versus actual percentage of students scoring 3 or 4 on the grade 8 science standardized test under each of the three orientations for school districts with at least one high school

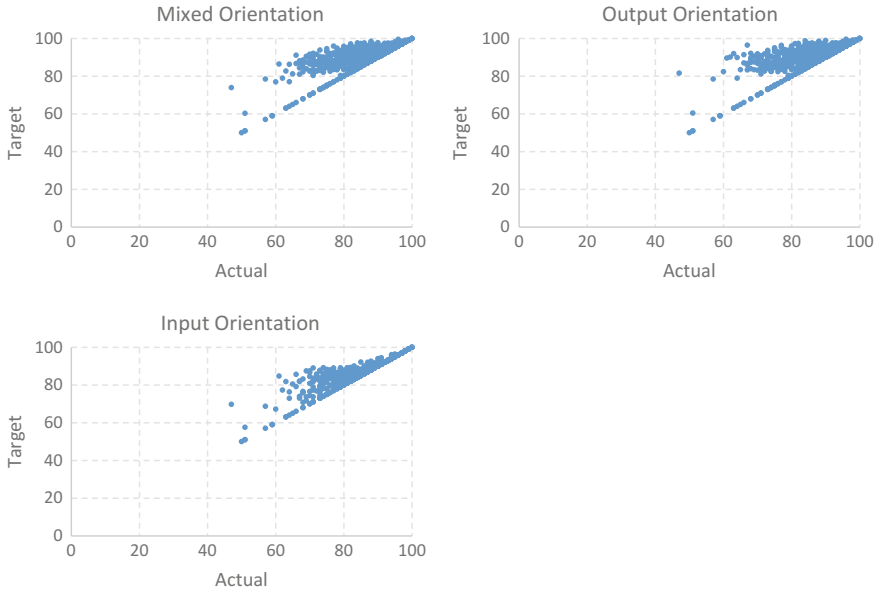


Fig. 19.9 Target versus actual percentage of 4-year graduation rate under each of the three orientations for school districts with at least one high school

Using a mixed orientation, we found evidence that the number of FTE teachers can be reduced by 0.2%, the number of FTE teacher support personnel by 4.3%, and the number of FTE building administration and professional staff personnel by 3.3%. In addition, the average³ percentage of students who score 3 or 4 on the English exam can rise by 0.4 percentage points, by 0.9 percentage points on the mathematics exam, and by 0.3 percentage points on the science exam.

Using a resource reduction orientation, we found evidence that the number of FTE teachers can be reduced by 0.8%, the number of FTE teacher support personnel by 4.6%, and the number of FTE building administration and professional staff personnel by 4.8%. In addition, the average percentage of students who score 3 or 4 on the English exam can rise by 0.6 percentage points, by 0.6 percentage points on the mathematics exam, and by 0.0 percentage points on the science exam.

Finally, using a performance enhancement orientation, we found evidence that the number of FTE teachers can be reduced by 0.0%, the number of FTE teacher support personnel by 4.3%, and the number of FTE building administration and professional staff personnel by 3.0%. In addition, the average percentage of students who score 3 or 4 on the English exam can rise by 0.4 percentage points, by 0.9 percentage points on the mathematics exam, and by 0.3 percentage points on the science exam.

³These are unweighted averages and therefore they do not represent the statewide percentages.

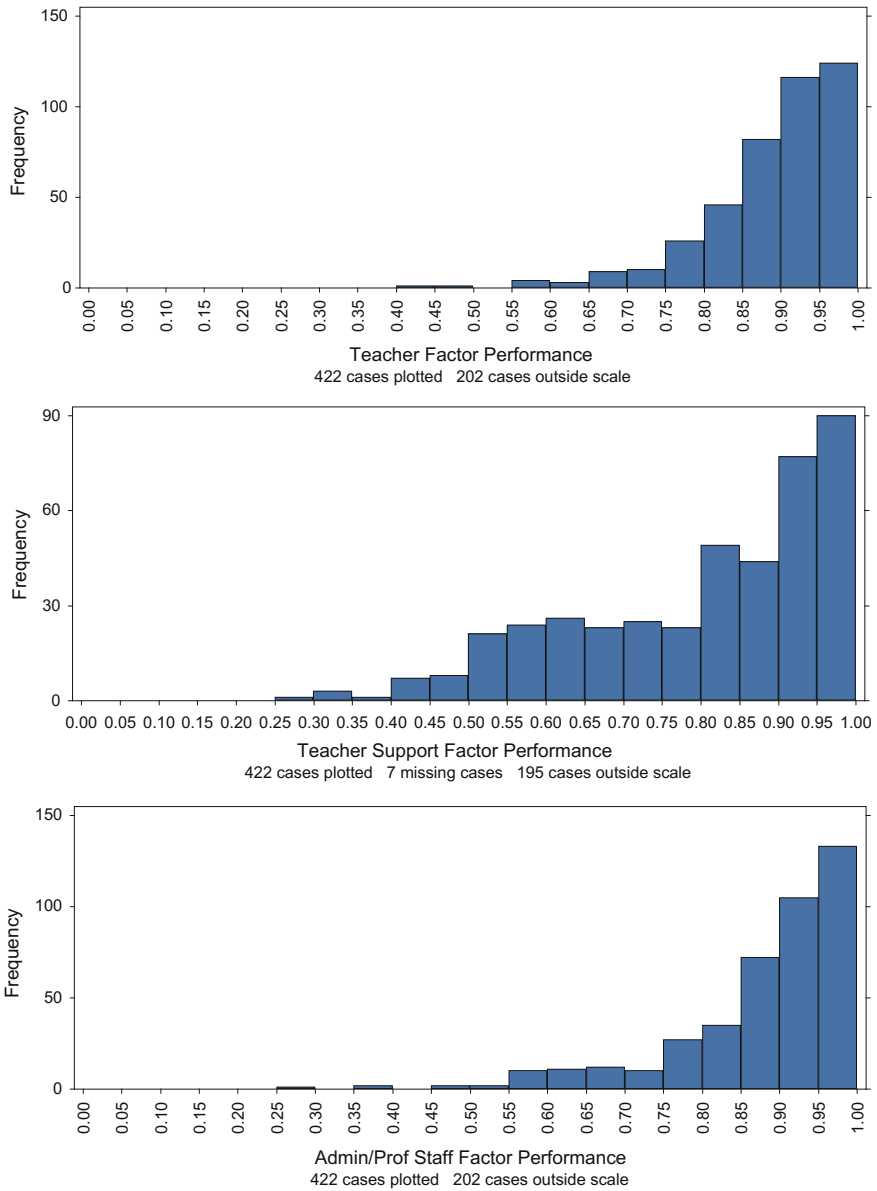


Fig. 19.10 Histograms of the school districts with at least one high school for each of the three factor performances associated with the resources, excluding those districts for which no improvement is possible

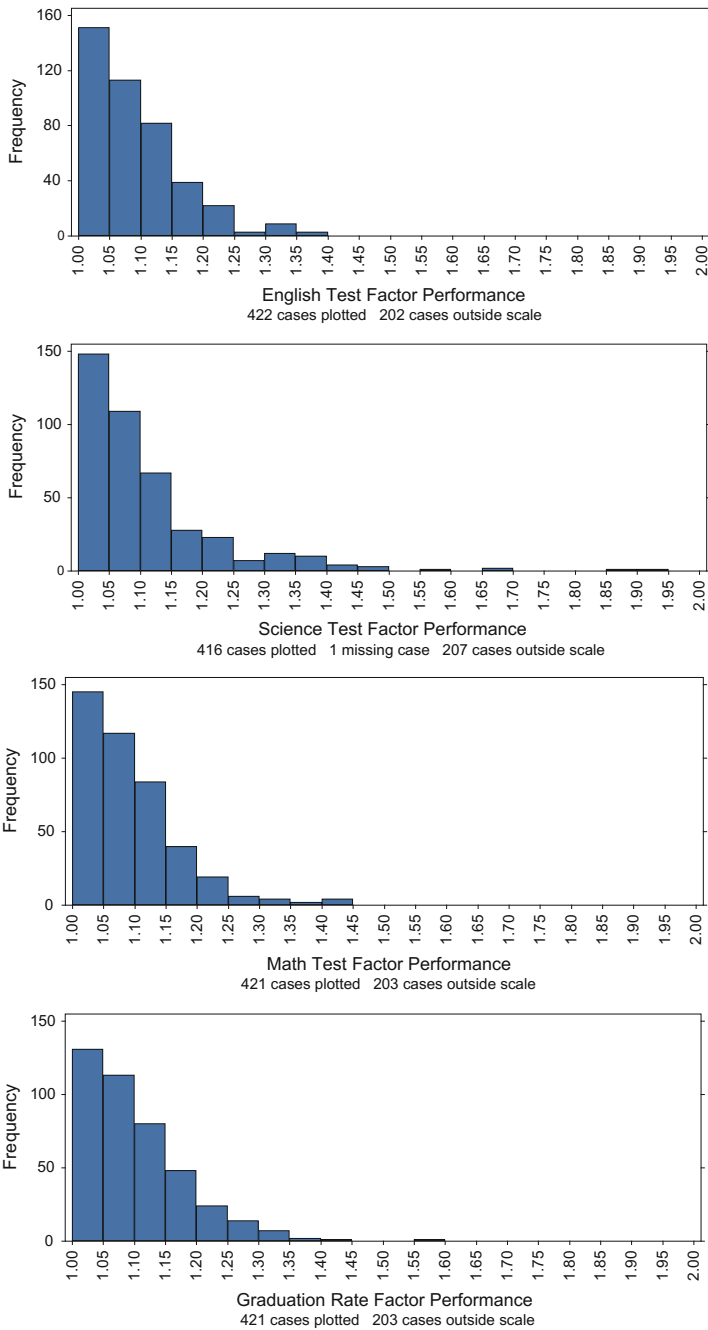


Fig. 19.11 Histograms of the school districts with at least one high school for each of the four factor performances associated with the performance measures, excluding those for which no improvement is possible

Table 19.3 Statewide results for all three orientations for School Districts without a high school

	FTE teachers	FTE teacher support	Building admin and prof staff	Grade 6 ELA (%)	Grade 6 math (%)	Grade 4 science (%)
Actual	2233	762	417	77.7	83.1	94.6
Mixed orientation						
Target	2228	729	403	78.0	83.8	94.8
Change	5	33	14	0.3	0.7	0.3
% Change	0.2	4.3	3.3	0.4	0.9	0.3
Resource reduction orientation						
Target	2216	727	397	78.2	83.6	94.6
Change	17	35	20	0.5	0.5	0.0
% Change	0.8	4.6	4.8	0.6	0.6	0.0
Performance enhancement orientation						
Target	2233	729	404	78.1	83.9	94.9
Change	–	33	13	0.3	0.7	0.3
% Change	0.0	4.3	3.0	0.4	0.9	0.3

The italicized values represent values in the direction of the model’s orientation

19.10 Implementation

We reiterate that other choices of variables are possible. An important first step is for the school districts and the New York State Education Department (NYSED) to work together to modify this model as necessary. For example, the current model does not include data on Regents exam scores. In principle, the only requirement is that complete data exists for all school districts for the specified school year. In addition, it is important to provide a complete data set so that all school districts, especially those in New York City, can be included. This data set needs to be compiled for the latest school year for which complete data are available.

The NYSED would need to determine the distribution of model results. Perhaps the initial distribution during a pilot phase should be restricted to the school districts and NYSED. This would allow school districts the opportunity to understand the full meaning of their own results better and to begin to incorporate the results into their operations and planning. The pilot phase would also allow school districts and NYSED to suggest further improvements in the model.

Ultimately, the model can serve as a key element in a quality improvement cycle. By providing direct feedback to each school district about its performance along multiple dimensions, it supports school district decisions about how to improve and allows them to demonstrate that their decisions have in fact had the desirable effects.

19.11 Intended Use of the Model

It is axiomatic that public school systems, being funded with public money, must be held accountable for the efficient use of their funding. They must use their available resources to provide the highest levels of learning possible given their myriad regulatory, faculty, student, and societal constraints—in effect, public school systems must solve a highly complex constrained optimization problem.

State and school system policymakers have not been able to hold school systems accountable for their performance in a way that precludes reasonable-sounding arguments about why any specific school district cannot perform at the performance level sought by policymakers. It has been too easy for school districts to claim that the target performance levels were set arbitrarily by unknowing policymakers, and that their school district has unique constraints that preclude it from meeting these arbitrary target performance levels. Of course, these claims could be true; we are not taking sides on this issue. We are only trying to solve it.

The model presented in this paper is designed and intended for the use of policymakers to effectively eliminate such claims. The model empowers policymakers to demonstrate that any school district not performing at the frontier level established by the model is underperforming. The school district(s) performing at frontier levels are doing so under conditions that are the same or worse than those school districts that are not performing at frontier levels.

Through the use of our model, policymakers will be empowered to establish data-informed performance levels for accountability, and school districts will not be able to avoid accountability by claiming uniqueness. School districts will then be forced to improve their efficiency or be identified openly and accurately as underperforming, in which case, the responsibility for underperforming will sit squarely with school district decision-makers. As those schools performing at frontier levels continue to innovate and improve, the frontier will continue to move outward, thereby eliminating any possibility of complacency within the accountability system by introducing healthy competition to remain on the frontier.

Our data-informed accountability model will incentivize the leaders of school districts not performing at frontier levels to improve their optimization of learning efforts or suffer being replaced with more effective leaders. This incentive has been missing within public school systems for decades and it will be transformative of public education if implemented.

19.12 Conclusions

We have presented a flexible model that allows school districts and NYSED to measure school district performance throughout New York State. The model provides multiple, mathematically derived performance measures that allow school districts to detect specific areas for improvement. The model also enables NYSED

to identify school districts that are the top performers in the state and others that most require improvement.

The results of a preliminary version of the model applied to data from the 2011 to 2012 school year shows that approximately one-third of the school districts in New York State are performing as well as can be expected given their local school district characteristics. Another 26.8–42.3%, depending on the specific resource or performance measure, can improve by no more than 10%.

Nonetheless, substantial statewide improvements are possible. Using the mixed orientation, for example, if every school district was to match to its target, New York State would have between 8 and 17% fewer personnel, 6 to 7% more students scoring 3 or 4 on standardized tests, and 6% more students graduating within 4 years.

Public education is critically important to the future of New York State and the nation. This model offers the potential to support public school education leaders in recognizing where improvements are possible and in taking appropriate action to implement those improvements.

Technical Appendix: The Mathematics of the DEA Model

We use two slightly different DEA models in this paper, one for school districts with one or more high schools, and one for school districts without a high school. The differences lie in the performance measures (different points at which test scores are measured, and no graduation rate for school districts with no high school). In addition, each model is employed with three different orientations (resource reduction, performance enhancement, and mixed). The text that follows describes the model for school districts with one or more high schools.

Let $n = 624$ be the number of school districts to be analyzed. The DEA literature refers to units under analysis as *decision-making units*, or DMUs. Let X_{ij} be amount of resource i consumed by DMU j , for $i = 1, 2, 3$, and $j = 1, 2, \dots, 624$. In particular, let X_{1j} be the FTE teachers in DMU j , let X_{2j} be the FTE teacher support in DMU j , and let X_{3j} be the FTE building administration and professional staff in DMU j .

Let Y_{rj} be performance measure r achieved by DMU j , for $r = 1, 2, 3, 4$ and $j = 1, 2, \dots, 624$. In particular, let Y_{1j} be the percentage of students scoring at levels 3 or 4 in secondary level English after 4 years of instruction in DMU j , let Y_{2j} be the percentage of students scoring at levels 3 or 4 in secondary level math after 4 years of instruction in DMU j , let Y_{3j} be the percentage of students scoring at levels 3 or 4 in Grade 8 Science in DMU j , and let Y_{4j} be the 4-year graduation rate as of August in DMU j , for $j = 1, 2, \dots, 624$.

Let S_{kj} be the value of site characteristic k at DMU j , for $k = 1, 2, 3, 4, 5$ and $j = 1, 2, \dots, 624$. In particular, let S_{1j} be the number of elementary school students in DMU j , let S_{2j} be the number of secondary school students in DMU j , let S_{3j} be the percentage of students with free or reduced price lunch in DMU j , let S_{4j} be the

percentage of students with limited English proficiency in DMU j , and let S_{5j} be the combined wealth ratio in DMU j , for $j = 1, 2, \dots, 624$.

The Resource Reduction DEA Model

The resource reduction DEA model with variable returns to scale, for DMU d , $d = 1, 2, \dots, 624$, is below. We must solve $n = 624$ linear programs to perform the entire DEA.

Min E_d	(1)	
Subject to		
$\sum_{j=1}^n \lambda_j X_{1j} \leq E_d X_{1d}$	(2.1)	FTE teachers
$\sum_{j=1}^n \lambda_j X_{2j} \leq E_d X_{2d}$	(2.2)	FTE teacher support
$\sum_{j=1}^n \lambda_j X_{3j} \leq E_d X_{3d}$	(2.3)	Building administration and professional staff
$\sum_{j=1}^n \lambda_j Y_{1j} \geq Y_{1d}$	(3.1)	Secondary level English (%)
$\sum_{j=1}^n \lambda_j Y_{2j} \geq Y_{2d}$	(3.2)	Secondary level math (%)
$\sum_{j=1}^n \lambda_j Y_{3j} \geq Y_{3d}$	(3.3)	Grade 8 science (%)
$\sum_{j=1}^n \lambda_j Y_{4j} \geq Y_{4d}$	(3.4)	Graduation rate (%)
$\sum_{j=1}^n \lambda_j S_{1j} \geq S_{1d}$	(4.1)	Number of elementary school students
$\sum_{j=1}^n \lambda_j S_{2j} \geq S_{2d}$	(4.2)	Number of secondary school students
$\sum_{j=1}^n \lambda_j S_{3j} \geq S_{3d}$	(4.3)	Percentage of students with free or reduced price lunch
$\sum_{j=1}^n \lambda_j S_{4j} \geq S_{4d}$	(4.4)	Percentage of students with limited English proficiency
$\sum_{j=1}^n \lambda_j S_{5j} \leq S_{5d}$	(4.5)	School district's combined wealth ratio
$\sum_{j=1}^n \lambda_j = 1$	(5)	Variable returns to scale
$\lambda_j \geq 0$ for $j = 1, 2, \dots, 624$	(6)	Nonnegativity
$E_d \geq 0$	(7)	Nonnegativity

We observe that setting $\lambda_d = 1$, $\lambda_j = 0$ for $j \neq d$, and $E_d = 1$ is a feasible, but not necessarily optimal, solution to the linear program for DMU d . This implies that E_d^* , the optimal value of E_d , must be less than or equal to 1. The optimal value, E_d^* , is the *overall efficiency* of DMU j . The left-hand sides of Eqs. (2)–(4) are weighted averages, because of Eq. (5), of the resources, performance measures, and site characteristics, respectively, of the 524 DMUs. At optimality, that is with the λ_j replaced by λ_j^* , we call the left-hand sides of Eqs. (2.1)–(4.5) the *target resources*, *target performance measures*, and *target site characteristics*, respectively, for DMU d .

Equations (2.1)–(2.3) imply that each target resource will be less than or equal to the actual level of that resource at DMU d . Similarly, Eqs. (3.1)–(3.4) imply that each target performance measure will be greater than or equal to the actual level of that performance measure at DMU d .

The nature of each site characteristic inequality in Eqs. (4.1)–(4.5) depends on the manner in which the site characteristic influences efficiency. Equations (4.1)–(4.4) correspond to unfavorable site characteristics (larger values imply a greater need for resources to obtain a given performance level, on average); therefore, we use the greater than or equal to sign. Equation (4.5) corresponds to a favorable site characteristic (larger values imply a lesser need for resources to obtain a given performance level, on average); therefore we use the less than or equal to sign. Thus, Eqs. (4.1)–(4.5) imply that the value of each target site characteristic will be the same as or worse than the actual value of that site characteristic at DMU d .

Thus, the optimal solution to the linear program for DMU d identifies a hypothetical target DMU d^* that, relative to DMU d , (a) consumes the same or less of every resource, (b) achieves the same or greater level of every performance measure, and (c) operates under the same or worse site characteristics. Moreover, the objective function expressed in Eq. (1) ensures that the target DMU d^* consumes resources levels that are reduced as much as possible in across-the-board percentage terms.

Of course, to proceed we must assume that a DMU could in fact operate exactly as does DMU d^* . In the theory of production, this is the assumption, made universally by economists, that the production possibility set is convex. In this context, the *production possibility set* is the set of all vectors $\{X_i, Y_r | S_k\}$ of resources, performance measures, and site characteristics such that it is possible for a DMU to use resource levels X_i to produce performance measures Y_r under site characteristics S_k . The convexity assumption assures that DMU d^* is feasible and that it is reasonable to expect that DMU d could modify its performance to match that of d^* .

We use the Premium Solver Pro[®] add-in (Frontline Systems, Inc., Incline Village, NV) in Microsoft Excel[®] to solve the linear programs. We use a macro written in Visual Basic for Applications[®] (VBA) to solve the 624 linear programs sequentially and save the results within the spreadsheet. Both the Basic Solver[®] and VBA[®] are available in all versions of Microsoft Excel[®]. However, the Basic Solver[®] is limited to 200 variables and 100 constraints, which limits the size of the problems to no more than 199 DMU and no more than 99 resources, performance measures, and site characteristics combined. We use the Premium Solver Pro[®], available from Frontline Systems, Inc., for this application.

The Performance Enhancement DEA Model

The performance enhancement DEA model with variable returns to scale, for DMU d , $d = 1, 2, \dots, 624$, is below. In this model, we eliminate E_d as the objective function (1) and from the resource constraints (2.1)–(2.3) and introduce θ_d as the new objective function (now to be maximized) and into the performance enhancement constraints (3.1)–(3.4). The parameter θ_d will now be greater than or equal to one, and it is called the *inverse efficiency* of DMU d .

Max θ_d	(1)	
Subject to		
$\sum_{j=1}^n \lambda_j X_{1j} \leq X_{1d}$	(2.1)	FTE teachers
$\sum_{j=1}^n \lambda_j X_{2j} \leq X_{2d}$	(2.2)	FTE teacher support
$\sum_{j=1}^n \lambda_j X_{3j} \leq X_{3d}$	(2.3)	Building administration and professional staff
$\sum_{j=1}^n \lambda_j Y_{1j} \geq \theta_d Y_{1d}$	(3.1)	Secondary level English (%)
$\sum_{j=1}^n \lambda_j Y_{2j} \geq \theta_d Y_{2d}$	(3.2)	Secondary level math (%)
$\sum_{j=1}^n \lambda_j Y_{3j} \geq \theta_d Y_{3d}$	(3.3)	Grade 8 science (%)
$\sum_{j=1}^n \lambda_j Y_{4j} \geq \theta_d Y_{4d}$	(3.4)	Graduation rate (%)
$\sum_{j=1}^n \lambda_j S_{1j} \geq S_{1d}$	(4.1)	Number of elementary school students
$\sum_{j=1}^n \lambda_j S_{2j} \geq S_{2d}$	(4.2)	Number of secondary school students
$\sum_{j=1}^n \lambda_j S_{3j} \geq S_{3d}$	(4.3)	Percentage of students with free or reduced price lunch
$\sum_{j=1}^n \lambda_j S_{4j} \geq S_{4d}$	(4.4)	Percentage of students with limited English proficiency
$\sum_{j=1}^n \lambda_j S_{5j} \leq S_{5d}$	(4.5)	School district's combined wealth ratio
$\sum_{j=1}^n \lambda_j = 1$	(5)	Variable returns to scale
$\lambda_j \geq 0$ for $j = 1, 2, \dots, 624$	(6)	Nonnegativity
$\theta_d \geq 0$	(7)	Nonnegativity

The Mixed DEA Model

The mixed DEA model with variable returns to scale, for DMU $d, d = 1, 2, \dots, 624$, is below. In this model, we keep both E_d and θ_d in the constraints and we may now choose to either minimize θ_d or maximize θ_d . We introduce a new constraint (6) that ensures balance between the goals of reducing resources and enhancing performance.

Min θ_d or Max θ_d	(1)	
Subject to		
$\sum_{j=1}^n \lambda_j X_{1j} \leq E_d X_{1d}$	(2.1)	FTE teachers
$\sum_{j=1}^n \lambda_j X_{2j} \leq E_d X_{2d}$	(2.2)	FTE teacher support
$\sum_{j=1}^n \lambda_j X_{3j} \leq E_d X_{3d}$	(2.3)	Building administration and professional staff
$\sum_{j=1}^n \lambda_j Y_{1j} \geq \theta_d Y_{1d}$	(3.1)	Secondary level English (%)
$\sum_{j=1}^n \lambda_j Y_{2j} \geq \theta_d Y_{2d}$	(3.2)	Secondary level math (%)
$\sum_{j=1}^n \lambda_j Y_{3j} \geq \theta_d Y_{3d}$	(3.3)	Grade 8 science (%)
$\sum_{j=1}^n \lambda_j Y_{4j} \geq \theta_d Y_{4d}$	(3.4)	Graduation rate (%)
$\sum_{j=1}^n \lambda_j S_{1j} \geq S_{1d}$	(4.1)	Number of elementary school students
$\sum_{j=1}^n \lambda_j S_{2j} \geq S_{2d}$	(4.2)	Number of secondary school students
$\sum_{j=1}^n \lambda_j S_{3j} \geq S_{3d}$	(4.3)	Percentage of students with free or reduced price lunch
$\sum_{j=1}^n \lambda_j S_{4j} \geq S_{4d}$	(4.4)	Percentage of students with limited English proficiency
$\sum_{j=1}^n \lambda_j S_{5j} \leq S_{5d}$	(4.5)	School district's combined wealth ratio
$\sum_{j=1}^n \lambda_j = 1$	(5)	Variable returns to scale
$E_d + \theta_d = 2$	(6)	Balance resource reduction and performance enhancement
$\lambda_j \geq 0$ for $j = 1, 2, \dots, 624$	(7)	Nonnegativity
$E_d, \theta_d \geq 0$	(8)	Nonnegativity

References

- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2, 429–444.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1979). Measuring the efficiency of decision making units: Short communication. *European Journal of Operational Research*, 3, 339.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1981). Evaluating program and managerial efficiency: An application of data envelopment analysis to program follow through. *Management Science*, 27, 668–697.
- Cooper, W. W., Seiford, L. M., & Tone, K. (1999). *Data envelopment analysis: A comprehensive text with models, applications, references and DEA-Solver software*. Boston: Kluwer.
- Driscoll, D., Halcoussis, D., & Svorny, S. (2003). School district size and student performance. *Economics of Education Review*, 22(2), 193–201.
- Emrouznejad, A. (2014). DEAZone. <http://www.deazone.com/>. Accessed 16 Sept 2014.
- Emrouznejad, A., Parker, B. R., & Tavares, G. (2008). Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. *Socio-Economic Planning Sciences*, 42, 151–157.
- Färe, R., Grosskopf, S., & Weber, W. L. (1989). Measuring school district performance. *Public Finance Review*, 17(4), 409–428.
- Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society Series A*, III, 253–290.
- Kang, B. G., & Greene, K. V. (2002). The effects of monitoring and competition on public education outputs: A stochastic frontier approach. *Public Finance Review*, 30(1), 3–26.
- Kirjavainen, T., & Loikkanen, H. A. (1998). Efficiency differences of Finnish senior secondary schools: An application of DEA and Tobit analysis. *Economics of Education Review*, 17(4), 377–394.
- New York State Department of Education. (n.d.). Guide to reorganization of school districts. http://www.p12.nysed.gov/mgtserv/sch_dist_org/GuideToReorganizationOfSchoolDistricts.htm. Accessed 17 Mar 2014.
- Ruggiero, J. (2007). Measuring the cost of meeting minimum educational standards: An application of data envelopment analysis. *Education Economics*, 15(1), 1–13.
- Sexton, T. R. (1986). The methodology of data envelopment analysis, In R. H. Silkman (Ed.), *Measuring efficiency: An assessment of data envelopment analysis*. New Directions for Program Evaluation No. 32 (pp. 7–29). San Francisco: Jossey-Bass.
- Sexton, T. R., Silkman, R. H., & Hogan, A. (1986). Data envelopment analysis: Critique and extensions, In Silkman, R. H. (Ed.), *Measuring efficiency: An assessment of data envelopment analysis*. New Directions for Program Evaluation No. 32 (pp. 73–105). San Francisco: Jossey-Bass.
- Thanassoulis, E., De Witte, K., Johnes, J., Johnes, G., Karagiannis, G., & Portela, C. S. (2016). Applications of data envelopment analysis in education. In *Data envelopment analysis* (pp. 367–438). New York: Springer.
- U.S. Census Bureau. (2011). *Public education finances*. <http://www2.census.gov/govs/school/11f33pub.pdf>. Accessed 17 Mar 2014.
- U.S. Government Spending. (n.d.). *New York state spending*. http://www.usgovernmentspending.com/New_York_state_spending.html. Accessed 17 Mar 2014.

Chapter 20

Efficiency Versus Effectiveness in Hospitals: A Dynamic Simulation Approach

Markus Schwaninger and Johann Klocker

Abstract Hospitals provide highly sophisticated services, but they are largely steered by means of simplistic management models, which do not match the complexities faced by these organizations. The design of management models in hospitals and public organizations at large shows a bend toward reductionism. The reductionism of these models is rooted in their short-termism, and in the myopia of their designers. The purpose of our contribution is to draft a path by which steering approaches can be developed, which are more effective in coping with organizational complexity than the short-termist, reductionist management models often in use. Using a generic model, we demonstrate that conventional approaches to steering entail unintended side effects leading to counterproductive system behaviors and to results inferior to those coming from no steering at all. We suggest how more sophisticated steering models can be designed to induce desirable modes of system behavior.

Keywords Health care · System dynamics · Mathematical modeling · Dynamic simulation · Case study

20.1 Introduction

Most services provided by hospitals are complex. They are sophisticated and call for high quality because health is of the essence in human life. The complexity of health care is a challenge to hospital management, for which effective management models are needed. What should they look like? At first it is easier to say what they are not. Frequently, hospitals are steered by means of simplistic management models

M. Schwaninger (✉)
University of St Gallen, St Gallen, Switzerland
e-mail: Markus.Schwaninger@Unisg.ch

J. Klocker
Landeskrankenhaus, Klagenfurt, Austria

(Grossmann and Scala 2002; Harper 2002), e.g., by one or a few budget figures that define

the total cost for one department. Such a figure is then divided by an output measure, e.g., the number of patient days. That ratio—cost per patient day—is deemed to express the efficiency of the hospital, and therefore is considered an appropriate device of control. We will show that it is not. Even if a few additional financial indicators are used, such models are reductionist and unsatisfactory.

What we can already claim at this stage is that to be effective a management model must be less simplistic, i.e., more sophisticated: It needs to go beyond the customary accounting figures, delve into the causal relationships underlying those figures, and provide a multidimensional view of the organization's development. And, most important, that view needs to provide a dynamic account of the system's behavior.

One might interject that these postulates are not new. Management models have been designed that fulfill some of these criteria. For example, the currently classical concept of the Balanced Scorecard (Kaplan and Norton 1996), which is used in a wide range of health organizations (Zelman et al. 2003), proposes steering a company on the basis of multiple indicators for performance measurement. The claim is that four dimensions of performance should be balanced in the management process: customers/ stakeholders, finance, internal business process, and organizational capacity. A Balanced Scorecard also depicts causal relationships between the indicators.

In a related vein, comprehensive systems of indicators have been developed (e.g., Gladen 2014). The respective indices, e.g., composite measures of financial performance (e.g., Carton and Hofer 2006), are organized in systematic forms, but in most of them little has been said about dynamic, causal relationships. The attribute "dynamic" in these studies is mostly related to growth rates of financial indices, not to changes of behavior in general.

In all of these cases, dynamic features such as feedbacks and delays are not systematically discussed. This is the main gap that we identify. We will try to propose paths to management models that account for the dynamic complexity that is characteristic of modern organizations. We will also suggest ways of accommodating both the *efficiency* and the *effectiveness* views in our concept of management models.¹ Finally, our approach concentrates on the domain of hospitals, which constitute a paragon instance of the intricacies faced by the organizations of our time. Of course, this contribution may be of relevant for both practitioners from any industry, and scholars, interested in the design of management systems.

Our study reverts to a case study from an Austrian hospital. We call it SANO. We use a single-case design because this enables revelatory insights. For that purpose, a simulation model was built as a didactic conceptual device. In order to

¹We define "efficiency" as the ratio between useful output and total input, or, more generally the ability of doing things well ("doing a task right"). In contrast, we define "effectiveness" as the degree to which a goal or desired condition is achieved ("doing the right task", also: doing something useful).

obtain a unit of analysis that enables transparency of causalities and meaningful results, we concentrate on one sector of the SANO hospital only, the oncology section. We start off by presenting the situation in that unit, and the issues it confronted at the outset. We then present a simulation model, which shows the causal structure between medical services, human resources, working climate, finance, and success in terms of both healed patients and economic results. The model is validated and different scenarios are simulated. Thereafter, different policies are examined and insights drawn from the simulations. The causal structures underlying success and failure are analyzed. Finally, we propose a path for the improvement of management models.

20.2 Case Study²

SANO is the central hospital of one of Austria's national states. The areas of health covered are comprehensive. The capacity is 2000 beds and the number of employees 4300, of which roughly 2200 (51%) are medical (600) and nursing (1600) staff.

Our focus in this study falls on the oncological care unit, and a pertinent network of medical services, which included several peripheral hospitals as well as local registered doctors from all over the state. We will call it Oncological Care System (OCS). In 2010, the OCS showed a record of successes, which had been achieved in its 25-year history. In 1985, when the unit was founded, the status quo of cancer treatment showed that the knowledge of oncology in the hospital was very unevenly distributed. Tumors were treated everywhere, often following the logic of the specialist in whose hand a patient had "landed", rather than by the optimal therapy from an oncological point of view. Therefore, the inspiration that guided the foundation of the oncology unit was a vision of oncological care based on an inter-, or even transdisciplinary³ way of providing highest level medical services.

This vision was put into practice in a rigorous effort by a core team of "champions" animated by very strong motivation and high professional ethos. Over the years an efficient system of oncological care was built for coping with the challenges ahead efficiently and effectively. Patients from the whole national state had quick access to treatment. The core of the oncology unit was located at SANO, providing treatment in a hospital ward and a tumor ambulance. But the oncological team also advised specialized departments, such as gynecology, urology, etc., in matters of oncology, by way of participating in their local ward rounds and by

²The second author (JK) was the medical doctor put in charge of building the Oncology Care Unit which is the object of this case study. The second author (MS) was an advisor to JK and built the simulation model under discussion in Sect. 20.3 of this chapter.

³In the present context, we conceive of "interdisciplinary" as a way of interacting among professionals from different disciplines. The attribute "transdisciplinary" then refers to a way of virtuous collaboration across disciplines, enabled through a shared theoretical framework or code.

gathering specialists from different specialized units in tumor boards. Finally, the oncologists provided decentralized services of that same kind to peripheral hospitals with little or no expertise in oncology. This had become more than a local unit for oncological care; it was worthy of being called an Oncological Care System (OCS). For details on its organization, see Schwaninger and Klocker (2017a).

The results achieved over the next few years were substantial. First of all, an organizational design was accomplished that made the patient its focal point, with all the features of a network organization. Patients and their families, often factored out of organizational plans, were the prime agents of the OCS. The central hospital and 9 more clinics, as well as registered doctors, were part of the care network, with a pivotal oncology unit as the main knowledge hub and coordinating agent. Among the innovative features of the structure were cross-sectional virtual teams, transdisciplinary collaboration, mobile units that brought doctors to the patient rather than the other way around, and networks both within and among the hospitals. The care process covered all phases from prevention to medical treatment to follow-up care and psychosocial accompaniment.

A second result was the remarkable performance of the OCS. Despite extremely scarce financial resources, both the quality and success of oncological care were increased. The system under study had become a showcase of holistic medical treatment that evoked sustained interest in professional circles all over Europe and beyond. The claim of remarkable performance has been scrutinized elsewhere (Schwaninger and Klocker 2017a). Based on data from the central Tumor Database, that study examined if there was any evidence of medicinal effects of the OCS between 1995 and 2013. Long-term data series on the evolution of 5-year survival rates across the state were used, for the five main entities of cancer indications (prostate cancer, lung cancer, etc.). 5-year survival rates are the most important indicator of effectiveness in oncological care (Ziegler et al. 2007). This analysis testified to success—a performance that could only be achieved through high quality of care.

Third, the OCS stood as an exemplar for the successful management of expertise. The influence of organizing and managing in general on the evolution of the system had become tangible. Yet it had not provoked the likely conflict between medical and managerial logics, which often impairs the qualifications of professionals (Boos and Mitterer 2014). Instead it brought to fruition a constructive force for the system's viability. The reason is twofold: On the one hand, management in this case never became a pathologically autopoietic⁴ system (Beer 1979). On the contrary, it was instrumental in pursuing the purpose of the OCS: a statewide, excellent level of care, enabled by transdisciplinary collaboration. The transdisciplinary nature of that collaboration was due to a shared framework for systemic management, and a common language which facilitated both daily operations and

⁴“Autopoietic” (from Greek) stands for the properties of self-production and self-maintenance of a biological or social system.

conversations about the future. For example, the distinction between efficiency and effectiveness, and the need to pursue both, were clear to all members of the staff.⁵

These outcomes exceeded all expectations. A major factor of success was the substantial freedom granted to the champions by the state health authority.

However, 25 years after the foundation of the oncological care unit, clouds appeared on the bright skies under which the OCS had sailed for such a long time. In 2011, the administration of the central hospital announced that it would cut the budgets of all departments, "... to improve the economic situation." The leaders of the OCS made a rough estimate: such a cut would reduce their financial resources available by roughly Euro 210,000 for the year. For a small unit this would be a painful cut. As the oncologists kept thinking about this, they discovered different implications that would profoundly affect the whole operation of the OCS.

20.3 Modeling, Qualitative and Quantitative

To analyze the situation more closely, a simulation model was built cooperatively: the oncologists, mainly the second author (JK), contributed the substantive knowledge about the issues under study, while the first author (MS) furnished modeling and simulation know-how. The model would be built off-site, i.e. outside the premises of the hospital, with the first author performing the technical modeling work. The purpose of the model was to anticipate what the implications of the announced cut in the budget would be. The users of the model would be the oncologists, but the results of the simulations were meant to be presented to the administrators of the hospital "at some point."

20.3.1 *Qualitative Model Components*

The dynamic hypothesis of the oncologists—to be formulated in the next section—was based on a qualitative analysis, which resulted in five causal loops, hereafter presented in Figs. 20.1, 20.2, 20.3, 20.4, and 20.5. The form we are using is that of Causal Loop Diagrams. Variables are connected by arrows indicating a causal relationship. The "+" signs on the arrows denote a reinforcing relationship: All arrows that carry a positive sign denote that the two connected variables point in the same direction, e.g., an increase of the causal variable induces an increase of the affected variable. The "-" sign indicates a balancing relationship. The arrows that

⁵We used *efficiency* more in the context of the short term, and *effectiveness* rather in terms of the long view.

Fig. 20.1 Finance–personnel loop

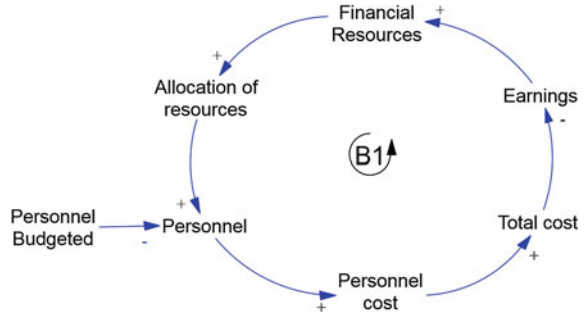


Fig. 20.2 Experience–stress loop

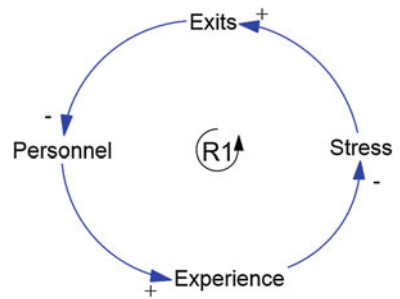
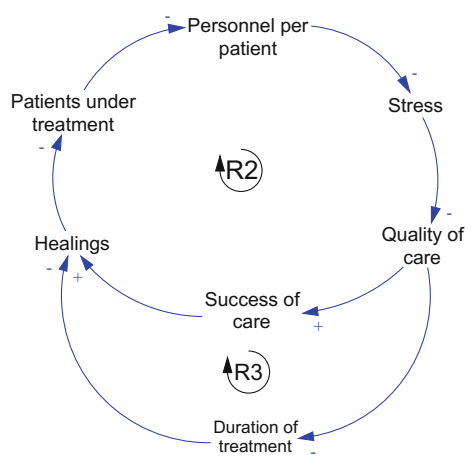


Fig. 20.3 Stress–quality loops



carry a negative sign denote that the two connected variables point in the opposite direction, e.g., that a rise in the causal variable diminishes the affected variable.⁶

⁶To make generally correct the statement “X and Y move in the same [opposite] direction”, a more precise formulation is necessary: “If X increases, Y increases above [below] what it would have been” (Richardson 1997).

Fig. 20.4 Goal adjustment loop

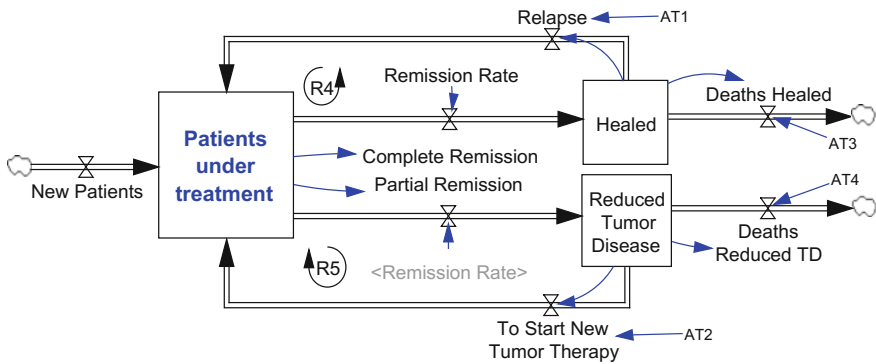
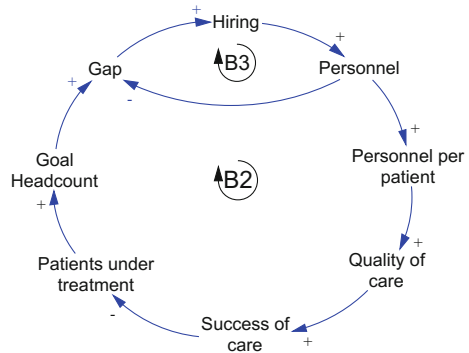


Fig. 20.5 Healing process loops

The Finance–Personnel Loop (Fig. 20.1) shows a connection that is straightforward: the allocation of financial resources enables hiring people therewith increasing the workforce. The larger the workforce, the higher the cost, which in turn decreases earnings and financial resources available. This is a balancing loop. Balancing loops are those loops that show a negative product of the signs (here: “+” * “+” * “+” * “+” * “+” * “-” = “-”), and are denoted with “B”. These loops are normally controlled by a goal or limiting factor. In this case, the personnel budgeted delimits the quantity of personnel.

Compared to the first loop, the second one (Fig. 20.2) highlights a counter-vailing relationship: the larger the workforce, the greater the experience and knowledge extant in the organization. The more experienced people are, the less susceptible to stress they become. Even so, stress is a proxy for a working climate that enhances the number of exits, which reduce personnel. This Experience–Stress Loop is self-reinforcing, leading either to a virtuous or a vicious cycle. Such

reinforcing loops show a positive product of all signs (here: “-” * “+” * “-” * “+” = “+”), and are denoted with “R”.

Figure 20.3 features two loops that show the causes and implications of both stress and quality of care. A lack of personnel leads to overload and stress, which is a major factor that jeopardizes quality of care. The success of care and the resulting number of cures alleviate the load of patients under treatment, consequently improving the personnel–patients ratio and alleviating stress. Lower stress means higher quality of care and then higher success of care. As the outer loop shows, quality of care—due to better dedication of staff and superior organization—reduces the duration of treatment, which affects the number of cures: shorter duration of treatment results in more healings. Both loops are of the reinforcing type.

The Goal Adjustment Loop in Fig. 20.4 again features a structure underlying a balancing behavior. The already discussed links from personnel to personnel intensity (personnel per patient) to quality of care, success of care (measured by remission rate), and patients under treatment, are brought together here with a crucial balancing factor, namely the adjustment to the goal for headcount (desired headcount). A staff target (“Goal Headcount”) higher than the actual headcount (“Personnel”) induces hiring activities, which enhance workforce size, etc. We may call this a “solution loop” because it would in theory lead to a better response to increasing patient loads.

For the next schema (Fig. 20.5) we have chosen the form of a Stock and Flow Diagram instead of a Causal Loop Diagram, in order to introduce the distinction between stocks and flows. This kind of diagram will then be used for the visualization of the quantitative simulation model.

The diagram is made up of two self-reinforcing loops that represent the Healing Process (Fig. 20.5). The variables in boxes are stocks that can either build up or be depleted. The variables pictured as valves are flows that increase or decrease the stocks. In this diagram most relationships between flows and stocks are of the “increase” type. The two loops demonstrate how the volume of patients under treatment builds up: in the upper loop (R4) more patients under treatment induce more healings (complete remission), triggered by a remission rate, i.e., the percentage of patients healed. That rate increases the number of healed persons (without tumor disease). The larger the stock of people without tumor disease, the more relapses there will be, and therewith more patients under treatment. All other aspects being equal, the number of patients under treatment tends to increase. This amounts to an Avalanche Effect.

If we were to expand our perspective, we would recognize that there are influences from outside: patients under treatment would grow also as a function of new patients, and the healed persons would decrease slowly as a function of deaths (see the complete model in Fig. 20.6). The lower loop (R5) replicates the logic of R4, for the fraction of partial healings.

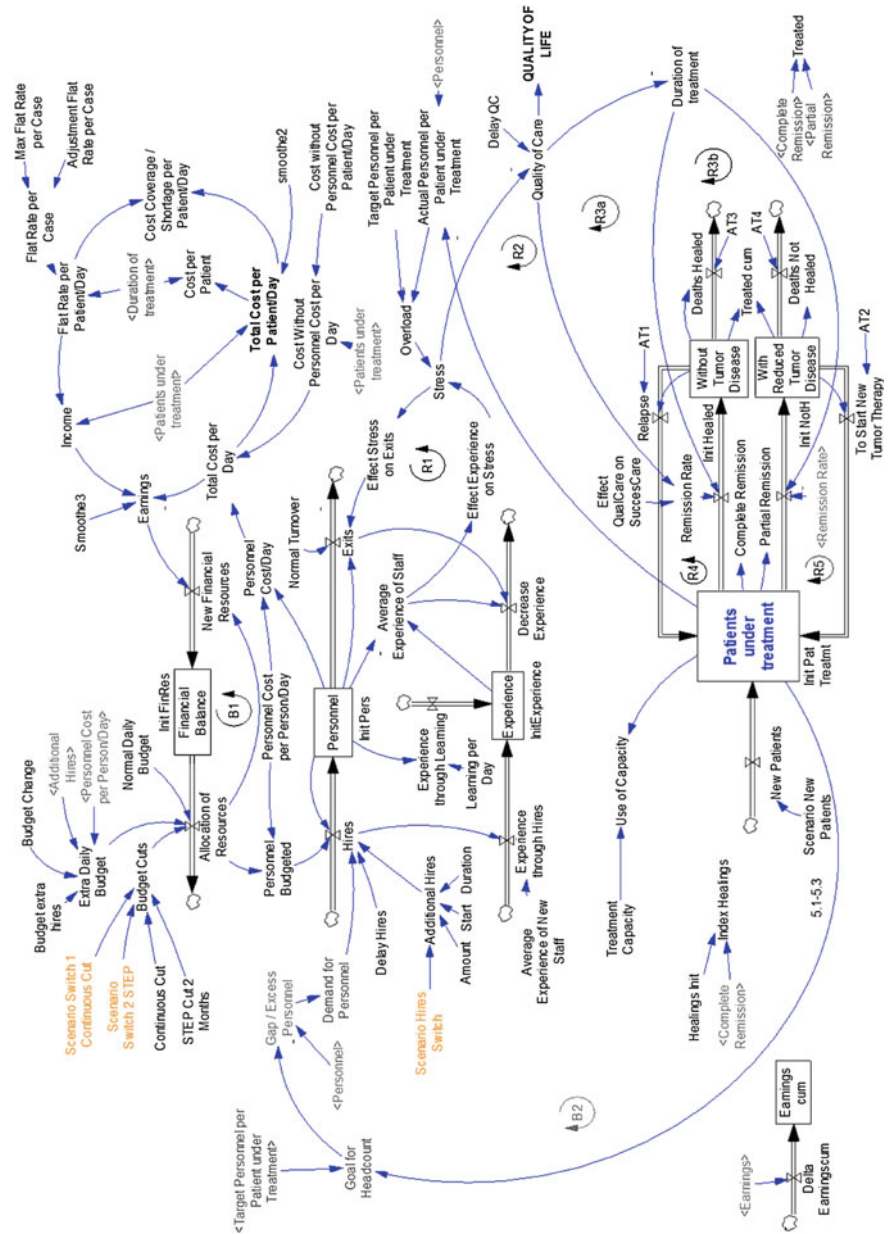
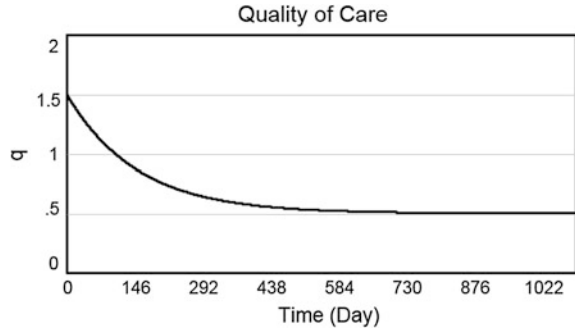


Fig. 20.6 Stock and flow diagram of the simulation model

Fig. 20.7 Dynamic hypothesis of an erosion of the quality of care



20.3.2 Dynamic Hypothesis

Based on these causal loops, the dynamic hypothesis of the oncologists was as follows: If the budget is cut, this will entail an increasing overload and stress of the personnel in the OCS and therewith lead to a decay in the quality of care, with a growing load of patients to be treated. The reference mode is shown in Fig. 20.7.

20.3.3 Quantitative Model

We have taken all the loops outlined above as building blocks and synthesized them into a quantitative simulation model. The overall picture of that model, in the form of a Stock and Flow Diagram, is shown in Fig. 20.6. We have abstained from introducing a stock for backlog, assuming that incoming patients had to be treated by all means.

The model is made up of equations, mostly differential equations. It covers a period of 3 years, from January 1, 2013 to December 31, 2015. It runs over 1095 days, with a small time step, sized 0.0625, to avoid rounding error.

To enable a full understanding of the model, the most important equations will be spelled out (dimensions of variables are in squared brackets):

20.3.3.1 Staff

The personnel headcount is the integral of hires and exits.

$$Personnel = Hires - Exits \quad [\sim \text{person}]. \quad (20.1)$$

The bottleneck for hires is the budgeted personnel. Additional hires exist only as a possibility in policy runs.

$$\begin{aligned}
 \text{Hires} = & ((\text{IF THEN ELSE}(\text{Demand for Personnel} < = \text{Personnel Budgeted} - \text{Personnel}, \text{Demand} \\
 & \text{for Personnel}, \text{Personnel Budgeted} - \text{Personnel}))) / \text{Delay Hires} + (\text{max}(0, \text{Personnel} \\
 & \text{Budgeted} - \text{Personnel}) + \text{Additional Hires}) / \text{Delay Hires} \quad [\sim \text{person/day}].
 \end{aligned}
 \tag{20.2}$$

Exits are a function of personnel turnover, with a multiplier for the effect of stress at work: the higher the stress the more exits.

$$\begin{aligned}
 \text{Exits} = & \text{Normal Turnover} * \text{Personnel} * \text{Effect Stress on Exits} \\
 & [\sim \text{person/day}].
 \end{aligned}
 \tag{20.3}$$

20.3.3.2 Services

The quality of care (QC), i.e. of the services, which are the product of the hospital, is essentially determined by the level of stress at work. A third-order delay denotes the inertia in the reaction of QC.

$$\text{DELAY3}(1/\text{Stress}, \text{Delay QC}), \text{Delay QC} = 30 \quad [\sim \text{day}]. \tag{20.4}$$

Stress is a function of Overload and triggered by a nonlinear multiplier that decreases with the level of the collective experience in the OCS.

$$\begin{aligned}
 \text{Stress} = & \text{Overload} * \text{Effect Experience on Stress}, \\
 & \text{Effect Experience on Stress} > 1 \quad [\sim \text{dimensionless}].
 \end{aligned}
 \tag{20.5}$$

Overload is the ratio of target over actual personnel per patient.

$$\begin{aligned}
 \text{Overload} = & \text{Target Personnel per Patient under Treatment} / \text{Actual Personnel per Patient under} \\
 & \text{Treatment} \quad [\sim \text{dimensionless}]
 \end{aligned}
 \tag{20.6}$$

20.3.3.3 Customers

Patients are a stock controlled by three inflows and two outflows.

$$\begin{aligned}
 \text{Patients under Treatment} = & \text{New Patients} + \text{Relapse} + \text{To Start New Tumor Therapy} - \text{Complete} \\
 & \text{Remission} - \text{Partial Remission} \quad [\sim \text{persons}].
 \end{aligned}
 \tag{20.7}$$

The use of capacity is a ratio of patients under treatment and capacity.

$$\text{Use of Capacity} = \text{Patients under treatment} / \text{Treatment Capacity} \quad (20.8)$$

[\sim dimensionless].

20.3.3.4 Economics

Earnings are determined by income and total cost per day.

$$\text{Earnings} = \text{Income} - \text{Total Cost per Day} \quad [\sim \text{Euro/day}] \quad (20.9)$$

The parameter values are documented in the Appendix, with dimensional units and sources. They are based on hard facts (*hf*) or estimates (*ed*) supplied by the doctors.

20.4 Simulations: Base Run

The simulation model was deployed, first to establish a number of scenarios that ascertained where the OCS was heading (“Base Scenario”), and later, after an account of model validation, what the consequences of the announced budget cut would be (“Budget Cuts”).

The Base Scenario pictures, *ex ante*, the path of the OCS if no changes would occur. In Fig. 20.8, the trajectories of 10 variables are plotted, based on the original, non-calibrated model. They show nonlinear patterns of behavior throughout.

If no changes occurred, the number of patients under treatment, and capacity use would grow along a sigmoid curve, with gradual changes in the first year. The use of capacity (not in the diagram), which is already loaded fully at the outset, becomes more and more stressed. The number of complete remissions shows strong growth as well, albeit along a decreasing slope. The dampening effect here is in the decrease of personnel. Why should the staff shrink? The personnel budgeted is constant. Hence, the answer lies in the exits, which are triggered by the growth of stress due to overload of the personnel. In addition, the greater turnover of staff, with a lag, leads to a loss of experience per staff member, starting after 19 months. Experience in its own right is a factor that dampens stress. As experience increases, stress first grows only moderately, but the ensuing institutional loss of experience leads to more growth of stress. As a consequence, the quality of care, which is the main indicator for patients’ quality of life, erodes markedly over the whole period. The duration of treatments increases continually, as a function of the drop in quality of care. Finally, one might think that the earnings would grow as the number of patients rises. However, as patients stay under treatment longer, the cost of care (cost per patient) increases as well, while the income stagnates. Hence, by month 17, the earnings move into a sustained negative status. Cumulated earnings become definitely negative in month 26.

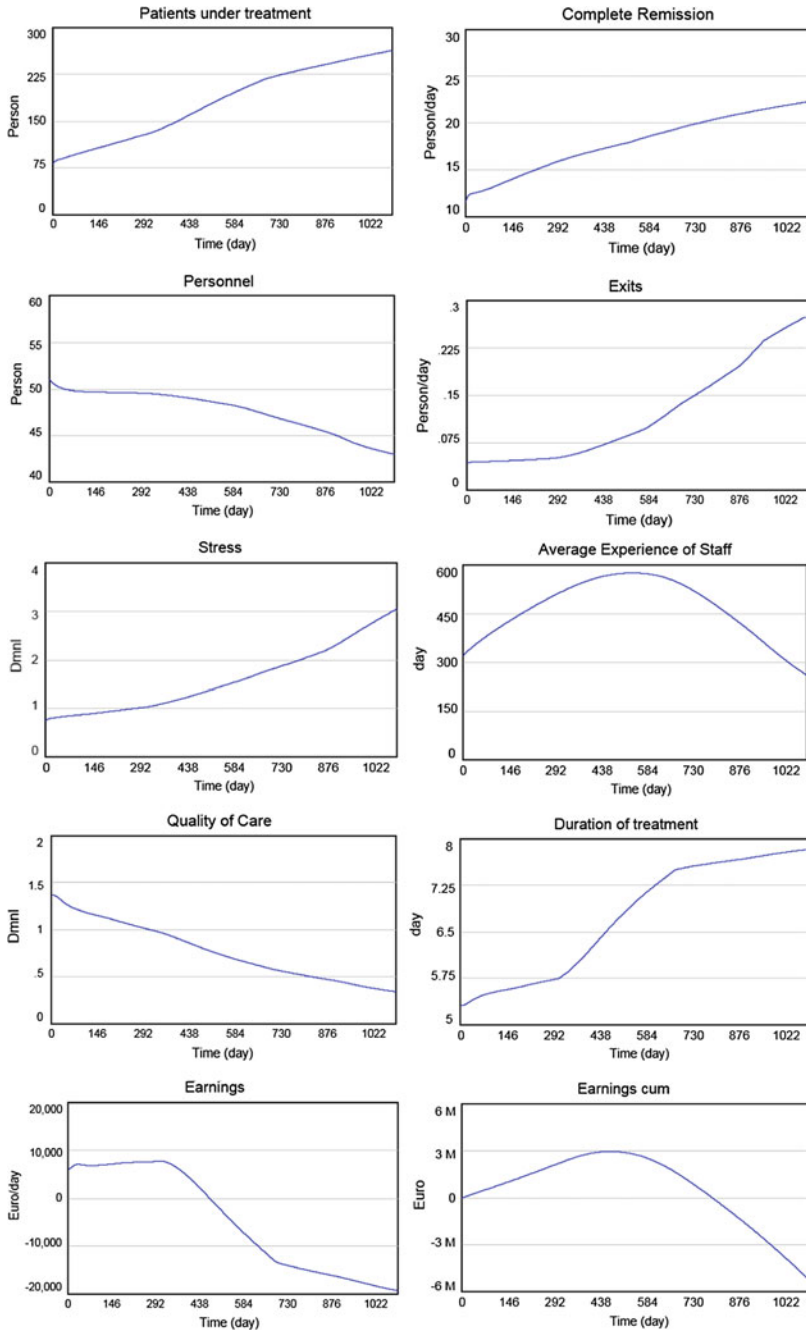


Fig. 20.8 Base scenario

Altogether, the OCS is in a situation of relative strength at the outset, but there are signals of a gradual weakening of that position, with some delay: Over the first 12 to 17 months, personnel, stress, and earnings change little, while experience even grows significantly. However, a less favorable situation is incipient, as stress builds up and the quality of care loses ground. Overload creeps into the system and causes stress, which jeopardizes the quality of care.

20.5 Model Validation

The processes of modeling, i.e., building the model, and validation, i.e., the effort to ensure quality of and confidence in the models, were simultaneous and intertwined.

The model was validated thoroughly in accordance with the principles and methods developed in the System Dynamics Community (Forrester and Senge 1980; Barlas and Carpenter 1996; Sterman 2000; Schwaninger and Grösser 2011). For the qualitative model, a dialogue took place between members of the OCS and the modeler, in the sense of structural validation, e.g., with structure and parameter examinations, as well as tests of boundary adequacy. All of these tests were integral to the modeling process, which only stopped once they were passed.

The first version of the quantitative model was built in 2012, for a first examination of the consequences of a budget cut. The results were presented at a gathering of the staff of the OCS from all over the State, with about 80 persons, including employees from all levels—leading medical executives to doctors and nurses. The contribution of that audience to validation was small, but important: they were given a rough walkthrough of the stock and flow diagram, and approved of its logical structure as well as the simulated behavior. So much for a first take on “face validity.”

In a nutshell, it became clear to the participants that a budget cut in the range of 10 to 20% would be counterproductive. At that stage, the administrators curtailed the budget by 15%.

The first version of the model showed the same features and produced roughly the same results as the second, calibrated version, even though it had not been submitted to a validation process as rigorous as the later version. That second version was carried out in early 2016, to enable an analysis in hindsight. It sought to determine if the model anticipated the development of the OCS adequately, and to what degree.

The validation of the quantitative model was essentially carried out by the first author as the external member and the second author as the internal member of the team. Most validation tests were of the indirect structural type, involving mainly sensitivity analyses, extreme conditions tests, dimensional consistency and integration error tests, all of which were passed. Loop knockout tests were realized for all reinforcing loops of the model. These tests resulted in certain plausible deviations of the focal variable (Personnel) as observed in the base run, but no implausible behavioral anomalies.

In addition, the mass balance check was performed (Dangerfield 2014). This is a powerful test, which has not been used much, due to its relative novelty. The procedure consists of accumulating all the inflows and outflows over time for each resource stock being modeled, and then using the following balance or checksum equation:

$$\int_{t=0}^{t=final} [\text{Sum of all inflows} - \text{Sum of all outflows} + \text{initial values of stocks} - \text{current values of stocks}] * dt = 0.$$

The correct result of the computation should be equivalent to zero throughout the run.⁷ If not, this would hint that the model was flawed. As with all other indirect structure tests, the model also stood this one.

Tests of model behavior were first carried out by qualitative expert judgment (“face validity”), in relation to all outcome variables, with numerous parameter variations. Surprise tests of variations in budget cut levels, patient inflow rates, and cost levels as well as stress- and experience-related indices did not show any previously unobserved behaviors.

As usual these validation tests involved several iterations, by which the clarity and performance of the model were improved. We did not carry out structural dominance analysis, which has a purpose similar to that of the loop knockout test—ascertaining abnormal behavioral changes as a consequence of changes in model structure. The former is more granular than the latter, and a test used mostly for large models (Oliva 2016). We deemed the loop knockout test sufficient in this case (see above). Family member tests were not performed either, but would be desirable in the future, in a possible endeavor of enhancing external validity.

In early 2016, as historical data became available, the model was submitted to behavior reproduction tests in the next round of validation experiments. For this purpose the two authors were supported by internal staff of the hospital, who provided essential data.

The quantitative model was calibrated on the accessible data series of personnel headcount and patient healings (complete and partial remissions). A list with the parameter values is available in the Appendix. In Fig. 20.9, the simulation outcomes of the calibrated model are compared with the historical data (“real”) obtained from the OCS.⁸

We are documenting the respective measures of fit in Table 20.1.

The correlation between simulated and historical data is close to the maximum of 100%. The Mean Squared Error, i.e. the averaged squares of the differences

⁷It is necessary to employ a double precision version of the software being utilized (in this case VENSIM).

⁸A cut of 15% of the budget is assumed (see scenarios in the next section).

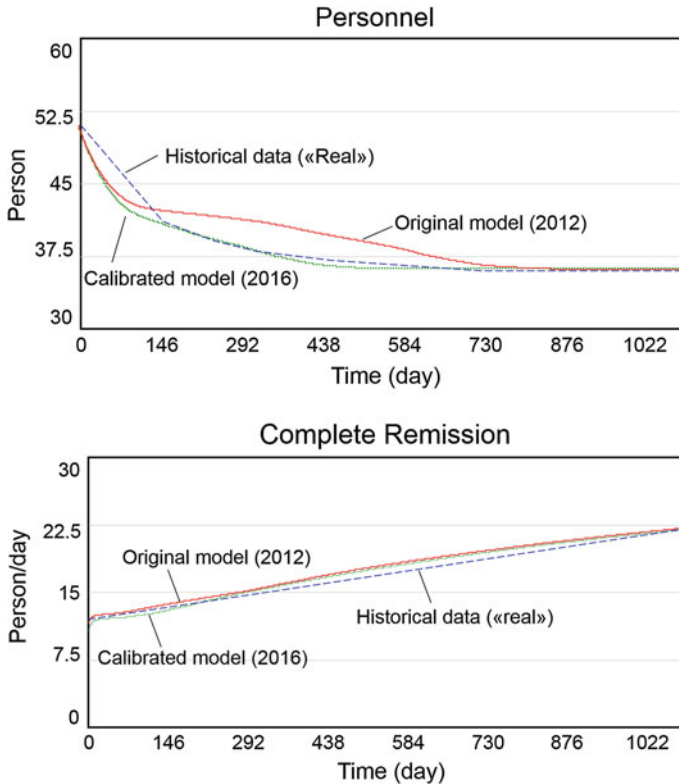


Fig. 20.9 Behavior reproduction of the model [Assumption: continuous budget cut of 15% (995 Euro/day)]

Table 20.1 Measures of fit with historical data of the simulations (Fig. 20.9)

Indicator	Size	Meaning
R^2	0.997425	Coefficient of determination
MSE	0.061587	Mean squared error
MAPE	0.520675	Mean absolute percentage error
U_m	1.41E-07	Mean difference
U_s	0.000491	Variation
U_c	0.999509	Covariation

between observed and simulated data, with 0.06, is very low. With a half percent the Mean Absolute Percentage Error is very small as well. The Theil Statistic explains the three components of the Mean Squared Error ($U_m + U_s + U_c = 1$). The simulation result shows no systematic bias caused by the differences in the dataset's mean value that would result in an overall upward or downward shift of the curve (U_m). The simulated curve closely depicts the variation of historical data with almost no difference in the magnitude of fluctuations (U_s). The differences between

the historical and the simulated datasets can be attributed largely to unsystematic point-by-point mismatch/discrepancy/noise (U_c).⁹ The analysis of the Mean Squared Error confirms that the model captures the overall trend in the historical behavior.

According to established standards the fit obtained is very high. Hence, strong confidence in the model is justified.

The scenario and policy runs to be accounted for in the next two sections are based on the calibrated model. In reality these simulations had to be realized before the calibration was possible—at the time when the original model came out (2012). Some of these runs on the original, non-calibrated model are recorded here in Fig. 20.8. More results of scenarios and policy simulations on the basis of the original model will be documented elsewhere (Schwaninger and Klocker 2017b).

20.6 Simulations: Budget Cuts

The Budget Cuts Scenario examines the implications of the measures announced by the hospital administration. This scenario assumes a continuous curtailment of daily budgets by 15% over the whole simulation period. We make the simplifying assumption that all budget retrenchments are applicable to staff expenses only, leading to decreases in the workforce.

The graphs in Fig. 20.10 show that the personnel drops enormously as a consequence of the budget cut.¹⁰ That leads to counterintuitive behavior of the system under study. The decrease of personnel intensity induces overload and markedly more stress, which accelerates the exit of employees, until month 16; then the exits plateau on a lower level than in the base run, due to the smaller volume of personnel. The precarious scarcity of employees induces a lower quality of care. The next consequence is a longer duration of treatments, i.e., people stay in hospital longer, so that the number of patients under treatment and therewith cost are enlarged. Another remarkable feature is that the number of complete remissions hardly decreases with the budget cuts (not in the graph). What recedes, and dramatically, is the quality of care—a cornerstone and lead indicator of a health care system: growing stress and falling staff experience occur at the price of unsatisfactory treatment of patients. This unexpected outcome makes visible a tradeoff whose importance cannot be overestimated: Quantity versus quality.

A second and even more surprising result concerns the economic dimension. The budget cuts appear to be successful in that fewer resources need to be allocated.

⁹For the uncertainty statistics used here see the original in Theil 1966, and the application to System Dynamics in Sterman 1984.

¹⁰That applies, as well, to temporary budget cuts for 1 or 2 years, but there the headcount rebounds, with a notable delay, as the budget goes back to normal.

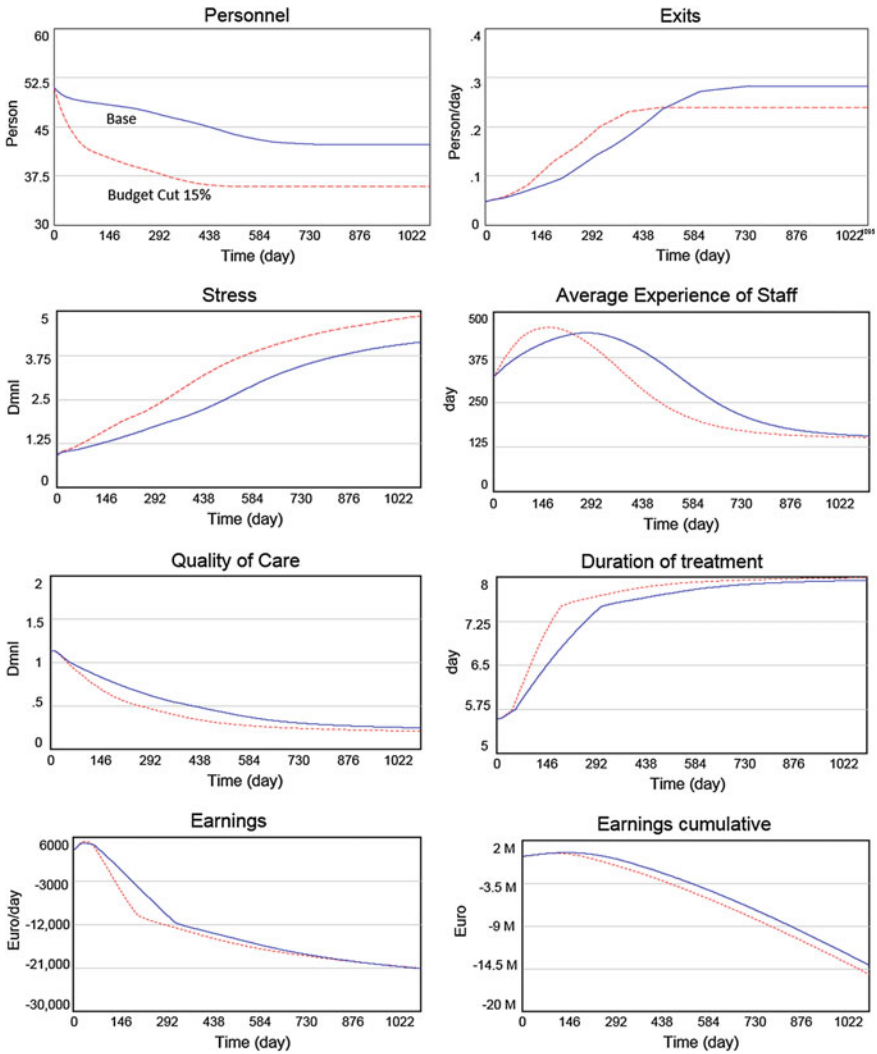


Fig. 20.10 Budget cut scenario (based on the calibrated model). *Solid line* base run (no budget cut); *dashed line* Budget Cut 15%, continuous

However, that impression is misleading. It turns out that the flow of earnings is even more negative in the budget cut scenario than in the base scenario. Correspondingly, the cumulative earnings become negative earlier than in the base scenario. In other words, not even the economic quantities respond to the interventions in a desirable way.

Much of the behavior of the model is counterintuitive from the viewpoint of the managers, while it makes sense from the stance of the medical staff. Even so, the

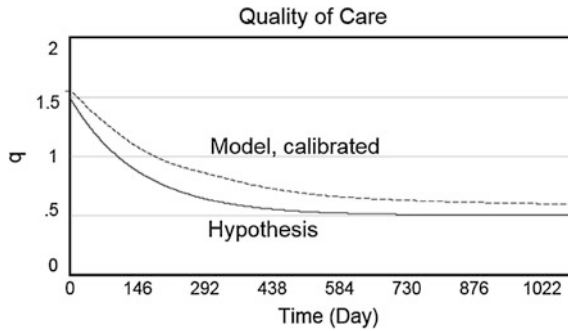


Fig. 20.11 Comparison of simulation outcome for personnel with reference mode (Assumption: Budget cut 15%)

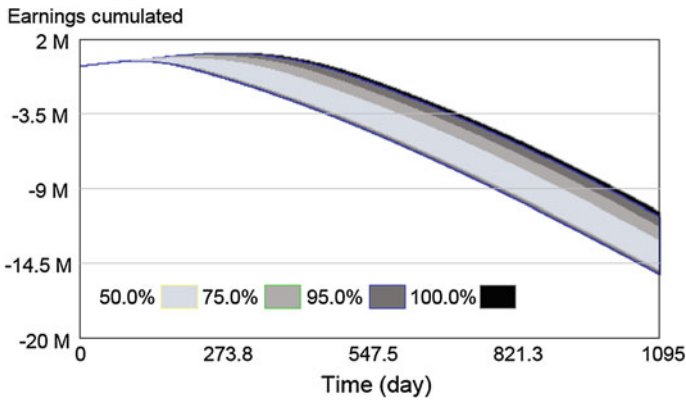


Fig. 20.12 Sensitivity analysis—impact of budget change on cumulative earnings (based on the calibrated model). 200 runs. The shades of *gray* show the results in bandwidths, each shade standing for a certain percentage of the total number of runs

working of the “mechanisms” just analyzed was fully understood by the doctors only in hindsight, when they saw the results of the simulations and had observable light bulb moments.

A comparison of simulation outcome with the assumptions made in the dynamic hypothesis, in line with the reference mode for quality of care (Sect. 20.3.2), is presented in Fig. 20.11.

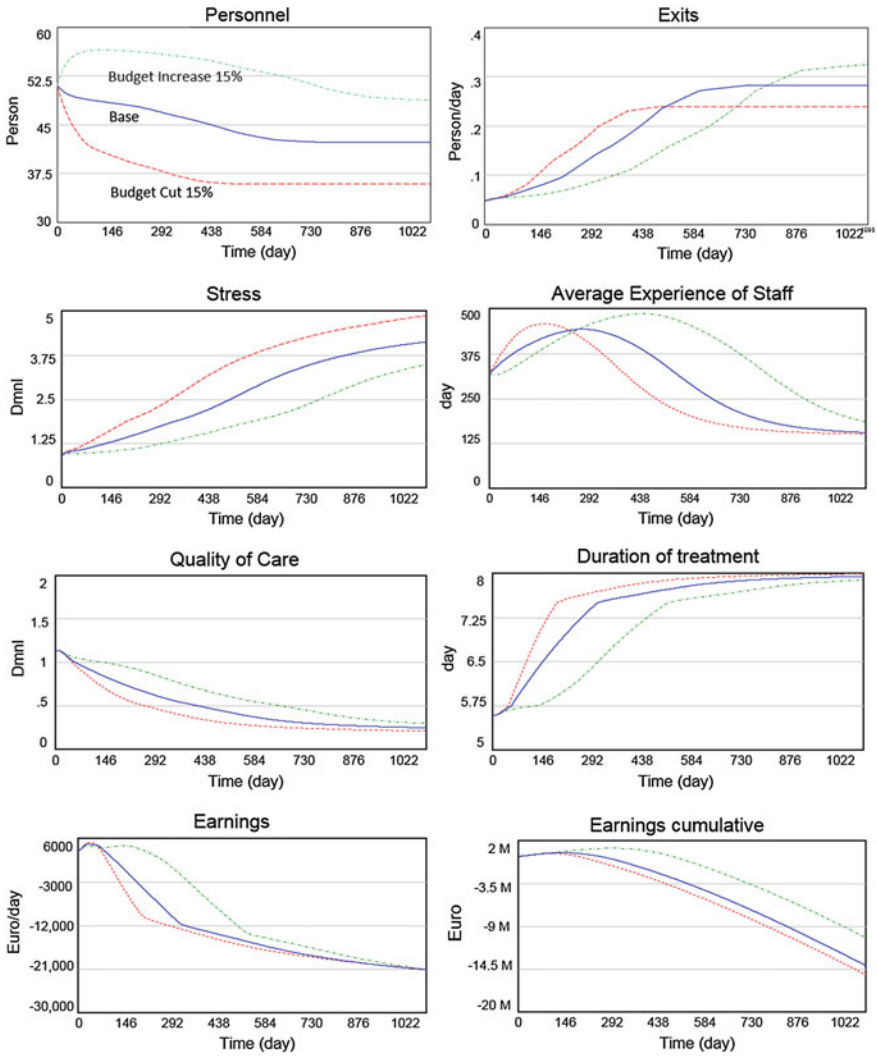


Fig. 20.13 Budget cut scenario (based on the calibrated model). *Solid line* Base run (no budget cut); *dashed line* budget Cut 15%; *spaced dashed line* budget increase 15%

20.7 Policies

Among possible sensitivity analyses we only present one, to answer the question: could changes in the daily budget bring about positive cumulative earnings? The sensitivity graph in Fig. 20.12 shows the results of 200 runs for a parameter variation of “Budget Change” from Euro -995 (i.e., -15%) to +995 (i.e., +15%) a

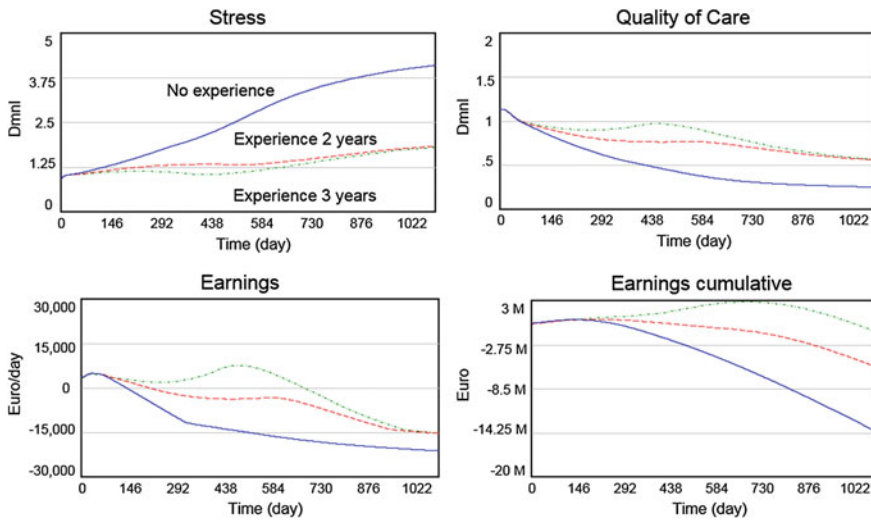


Fig. 20.14 Sensitivity analysis—impact of experience of persons hired (based on the calibrated model). *Solid line* No experience; *dashed line* experience 2 years; *spaced dashed line* experience 3 years

day. The cumulative earnings are plotted with the shades of gray symbolizing the different confidence levels, i.e., the respective percentages of the amount of total runs. A closer analysis of these results shows that cumulative earnings stay positive for 8.5 months, then turn negative. With a budget cut of 15% they already become negative after 6 months. On the other hand, an expansion of the budget by 15% leads to positive cumulative earnings for 16 months.

The interesting question now is what impact a budget increase would have on other variables. In other words, what would be the implications of such a policy? In Fig. 20.13 we show a comparison of all three runs—base, budget cut (15%), and budget increase (15%).

The dot and dash lines mark the budget increase scenario. The projected increase would mitigate personnel scarcity and consequently stimulate the whole system under study to much higher performance: stress would decrease substantially, while the staff would become more experienced, and the quality of care would grow. As the duration of treatments would be abridged, the financial situation would improve. It must be noted that the budget increase scenario would be contingent on support from the public sector.

Another policy domain examined is hiring. Our sensitivity analysis shows that hiring people with more professional experience entails positive consequences throughout: in the case of an experience level of zero years of staff newly hired, the consequences are drastic throughout (Fig. 20.14). In comparison, for an experience level of 2 or 3 years, stress remains stable, and the decay in quality of care can be mitigated significantly. For an experience level of 3 years, earnings stay positive for almost 2 years (in the zero experience scenario for 4.5 months only),

and cumulative earnings remain above zero, except for the last 2.5 months (Fig. 20.14), etc. The advantages of a policy of hiring experienced people are obvious and impressive. However, it is very difficult to implement it, as the market for hospital staff in the region has dried out.

Further scenarios were run to explore different policies, namely changes in resource allocation. In addition, scenarios and sensitivity analyses for showing the implications of different levels of inflows of new patients were elaborated. Policies to mitigate capacity shortages were examined as well. We abstain from presenting a detailed log of these additional simulations undertaken, because they transcend the purpose of this chapter.

The scenarios referred to here lead to a first conclusion transcending the specific case of the OCS. Reflecting on the results obtained, we propose that the patterns observed apply to any health organization. In addition, it can be assumed that the features of the model presented here are of a generic type. We surmise that the structure of our model could be an initial paragon for the elaboration of similar models in a broader range of organizations, namely of the services sectors.

20.8 Discussion and Implications

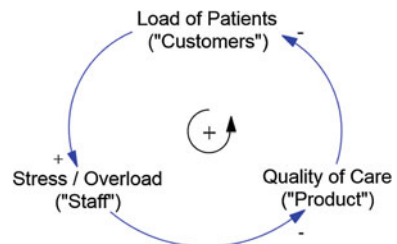
Our simulations support the dynamic hypothesis formulated earlier: the assumption was that a budget cut would lead to an increasing overload and stress of the OCS personnel and therewith to a decay in the quality of care, with a growing load of patients to be treated.

The simulations confirm all three aspects inherent in this hypothesis and, additionally, show that three tendencies reinforce each other mutually. This is shown in the high-level causal loop diagram in Fig. 20.15.

This diagram captures relationships that are in play among three perspectives—customers, staff and product. This high-level causal structure reveals a “mechanism” by which success and failure are brought about.

The example of the small model developed and explored here shows how a multidimensional, dynamic account of complex system behavior can be provided as a management support. It has also been shown that such a model is superior to conventional management systems.

Fig. 20.15 High-level diagram of three mutually reinforcing factors leading to crisis



The model presented amounts to more than a photography of a system state. It captures the dynamics of both the short run, with the efficiency view, and the long term, with an effectiveness perspective. If the budget is cut, the consequence is a decrease in personnel costs; this is a short-term success. However, in the longer run, antagonistic forces emerge, not only in economic terms. More important, the purpose of the hospital—quality of care, together with the healings quota—is affected. This difference between the short-term and long-term views is a clearly palpable instance of the distinction between efficiency and effectiveness.

It is not necessary to increase the budget, but it is imperative not to cut resources painfully. While budget cutting offers the path of least resistance, promising the relief of economic concerns, that short-termist attitude is refuted by the longer term consequences: it paves the way to disaster. As quality erosion creeps in, the virtuous path of success is lost. Vicious circles establish themselves, even before their consequences become palpable. Once these vicious loops are established, there is hardly any path back to a virtuous trajectory.

To give this discussion continuity, we will propose a path for the improvement of management models.

20.9 Toward the Improvement of Management Through Models

There are two crucial aspects for the improvement of a management model. One concerns the structure of the model to be developed, while the other shows the way to proceed in its design and implementation.

The substantive aspect derives from the need for better models as outlined early in this article. The conventional models, tending to be simplistic and reductionist, epitomize a danger for the governance of an organization. Therefore, new paths must be found to enable virtuous and effective management. First of all, a management model needs to be multidimensional. This can even be achieved with a small model like the one we have presented. Instead of relying on one or a few economic indicators, we have implemented indicators covering four dimensions, related to customer, product, staff, and economics. Second, such a model also needs to be dynamic, and the model presented here is a dynamic one. The dynamics of the system are generated by feedbacks and delays. They reflect a dynamic behavior mainly produced by the relationships among endogenous variables¹¹. External variables play only a minor role. This highlights the nature of systemic problems: they are “homemade” and must not be relegated to exogenous influences. The advantage of this kind of management system is that it allows one to look at the

¹¹The explanation of a system’s behavior by endogenous structure has been highlighted by Forrester (1968) and Richardson (2011) as a main feature of the system dynamics methodology.

state of affairs under the surface, tracking causal relationships that are normally invisible.

The second aspect is about process. In the case under study, the model and the simulation results were presented to the staff of the oncological unit, who agreed with our analysis and appreciated the insights gained from it. Subsequently, however, it turned out that the impact on the governance of the unit was nonexistent. One imputed reason could be that the management that went about cutting the budget was unreasonable, stubborn, and so on.

However, we know from research in the behavioral sciences that people are motivated by a sense of “psychological ownership,” a feeling that they can lay claim to certain organizational factors as their own (Pierce et al. 2001). Such ownership fosters self-identification as well as organizational commitment and citizenship behavior (Van Dyne and Pierce 2004). A major drawback is that the managers who initiated the budget cut had in no way been involved in the formation of the simulation model. That model had been developed rather spontaneously between the two authors. They did not consider, at that point, what in hindsight appears as a straightforward way to proceed: trying to involve the administrators in the construction of the model and using the model in the political negotiation process.

In other words, the administrators were not acquainted with the rationale developed by the doctors who were the frontline owners of that model. A much better approach for proceeding would have been to involve those managers in the discussion around the consequences of the budget cut and the construction of the model. This could have changed their mental models and prompted a different outlook on the issues at hand, giving them a grasp of the dynamics generated by their decision. The involvement in the model building exercise would have given them a feeling of ownership and maybe led to a different decision, for example, a decision of undertaking no budget cut at all (Vennix 1996).

Our proposition then is:

The improvement of a management system can be best obtained not only by designing a dynamic, multidimensional model, but also giving decision-makers a sense of ownership of the system under development through their participation in the design process.

These are necessary conditions without any claim of completeness. We have, for example, not delved into the question of matching the management model with the other management systems, and with other components of the management framework, e.g., organizational culture, structure, and strategy (Schwaninger 1994). Such reconciliation is another crucial factor for improving the effectiveness of management by means of models of the kind presented here.

20.10 Conclusions

We have explored a path by which the reductionism in much of the established management systems can be overcome. In line with the purpose of our study we have given some pointers about the limitations of short-termist, reductionist management models. Subsequently, reverting to modeling and simulation, we have drafted a path by which multidimensional, dynamic models can be built and used to support leaders and managers more effectively in their coping with complexity.

At the substantive level, as manifest in the model presented here, our study uncovers a structure that generates characteristic patterns of behavior. These conform to the expectations of the medical staff, but are counterintuitive in the logic of the administrators (Forrester 1971). *Our main finding of that kind—detected ex ante and corroborated a posteriori—is that the intervention of a budget cut, contrary to the expectations of the administrators, led to a decrease in earnings.* This is probably a feature that can be found in many organizations,—an archetypal structure so to say (Senge 1990; Wolstenholme 2003).

In that sense, we can assume that the structural features of our model are of a generic type. In other words, they are applicable to multiple contexts, representing a “wider class” of real-world situations (Forrester 1961, p. 208). Apparently these features exist in a broad range of organizations. The patterns of behavior ascertained here, and their underlying structures, can be found not only in hospitals but in any kind of service organization, e.g., many public institutions. Beyond that, according to our general experience, they are also relevant for private firms.

More effective steering requires more sophisticated management models. That is an insight in line with the Conant-Ashby theorem: “Every good regulator of a system must be a model of that system.” (Conant and Ashby 1981). In other words, the result of a process, e.g., a management process, can only be as good as the underlying model. That is, the quality of the models used is crucial for managerial effectiveness. This article presents an application of the system dynamics methodology as an effective device for the improvement of management models. It could therefore be a useful conceptual input for those who strive to make better decisions. At the same time our study should also be of interest to researchers dealing with the design of management systems, in hospitals and beyond.

At this point we also have to reflect upon the limitations of our study. First, the simulation model presented here is not complete; it does not cover all relevant variables and relationships concerning the system under study. For example, the linkages between work experience, stress, and quality of care—represented in the model—cover a fundamental qualitative feature of organizational culture. Related aspects such as systematic psychohygienic measures for the staff, and sustained efforts of the management to strengthen culture, are prominent in the OCS. We have refrained from quantifying these latter aspects, on purpose. However, we underline that these factors also impinge on the trajectory of the OCS, and might even be the main factor warranting its resilience in this time of turbulence.

Second, management models are not a panacea against all threats and weaknesses menacing an organization. They cannot be better than the management in whose hands they rest. But the problem starts earlier:

Third, many leaders and managers are not aware of the virtues of dynamic models, and refuse them with the verdict “too complicated!” These leaders are simply unaccustomed to working with such simulations: thinking in loops, delays, and dynamic behavior is not a common practice, which means that it is not easy to “sell” modeling and simulation as something beneficial for management.

Finally, there are also certain technical barriers to the use of these models: the construction and use of good models require certain modeling skills which must either be built into the organization or acquired from outside.

We trust that these limitations are not overwhelming. Dynamic modeling and simulation are of growing importance for the quality of management. They are already becoming a force for supporting decision-making processes in new and promising ways.

Acknowledgements The authors wish to thank two anonymous reviewers for their insightful critiques. Special thanks go to Ms. Evgenia Ushakova for her valuable comments on our text and precious help in technical matters. The authors are also grateful to Dr. John Peck for his excellent review of the English.

Appendix: Parameter Values

Parameter	Value in the original model	Unit	Source	Value in the calibrated model
Normal Daily Budget	6630	Euro/day	ed	6562
Initial Personnel	51	Person	hf	51
Personnel Cost	130	Euro/(day*person)	hf	130
Delay Hires	60	Day	ed	58
Initial Experience	16320	Day*person	ed	16,256
Normal Personnel Turnover	0.0033	1/day	ed	0.0033
Cost without Personnel Cost per Patient/Day	500	Euro/person/day	ed	500
Maximal Flat Rate per Case	3500	Euro/person	ed	3500
Treatment Capacity	80	Person	hf	80
Delay Quality of Care	30	Day	ed	35
Target Personnel per Patient under Treatment	0.625	Dimensionless	ed	0.75

Code: ed estimate doctors; hf hard facts

References

- Barlas, Y. (1996). Formal aspects of model validity and validation in system dynamics. *System Dynamics Review*, 12(3), 183–210.
- Beer, S. (1979). *Heart of enterprise*. Chichester: Wiley.
- Boos, F., & Mitterer, G. (2014). *Einführung in das systemische Management*. Heidelberg: Carl-Auer-Systeme Verlag.
- Carton, R. B., & Hofer, C. W. (2006). *Measuring organizational performance: Metrics for entrepreneurship and strategic management research*. Cheltenham, UK: Edward Elgar.
- Conant, R. C., & Ashby, W. R. (1981, originally published in 1970). Every good regulator of a system must be a model of that system. mechanisms of intelligence: Ross Ashby's writings on cybernetics (pp. 205–214) (R. C. Conant, Ed.). Seaside, CA: Intersystems Publications.
- Dangerfield, B. (2014). Systems thinking and system dynamics: A primer. In S. Brailsford, L. Churilov, & B. Dangerfield (Eds.), *Discrete-event simulation and system dynamics for management decision making* (pp. 26–51). Chichester: Wiley.
- Forrester, J. W. (1961). *Industrial dynamics*. Cambridge, MA: MIT Press.
- Forrester, J. W. (1968). *Principles of systems*. Waltham, MA: Pegasus Communications.
- Forrester, J. W. (1971). Counterintuitive behavior of social systems. *Technology Review*, 73(3), 52–68.
- Forrester, J. W., & Senge, P. M. (1980). Tests for building confidence in system dynamics models. In A. A. J. Legasto, J. W. Forrester, & J. M. Lyneis (Eds.), *system dynamics* (pp. 209–228). Amsterdam: North-Holland.
- Gladen, W. (2014). *Performance measurement controlling mit Kennzahlen*. Wiesbaden: Springer Fachmedien Wiesbaden.
- Grossmann, R., & Scala, K. (2002). *Intelligentes Krankenhaus: Innovative Beispiele der Organisationsentwicklung*. Wien/New York: Springer.
- Harper, P. R. (2002). A framework for operational modelling of hospital resources. *Health Care Management Science*, 5, 165–173.
- Kaplan, R. S., & Norton, D. P. (1996). Using the balanced scorecard as a strategic management system. *Harvard Business Review*, 74(1), 1–14.
- Oliva, R. (2016). Structural dominance analysis of large and stochastic models. *System Dynamics Review*, 32(1), 26–51.
- Pierce, J. L., Kostova, T., & Dirks, K. (2001). Toward a theory of psychological ownership in organizations. *Academy of Management Review*, 26, 298–310.
- Richardson, G. (1997). Problems in causal loop diagrams revisited. *System Dynamics Review*, 13(3), 247–252.
- Richardson, G. P. (2011). Reflections on the foundations of system dynamics. *System Dynamics Review*, 27(3), 219–243.
- Schwanning, M. (1994). *Managementsysteme*. Frankfurt, New York: Campus.
- Schwanning, M., & Grösser, S. (2011). System dynamics modeling: Validation for quality assurance. In R. A. Meyers (Ed.), *Complex systems in finance and econometrics* (Vol. 2, pp. 767–78). New York: Springer.
- Schwanning, M., & Klocker, J. (2017a). Holistic systems design: The oncology carinthia study. In C. García-Díaz & C. Olaya (Eds.), *Social systems engineering*. Chichester: Wiley.
- Schwanning, M., & Klocker, J. (2017b). Systemic development of health organizations: An integrative systems methodology. In H. Qudrat-Ullah & P. Tzasis (Eds.), *Innovative health systems for the 21st century*. New York: Springer.
- Senge, P. M. (1990). *The fifth discipline: The art and practice of the learning organization*. New York: Currency Doubleday.
- Sterman, J. D. (2000). *Business dynamics. Systems thinking and modeling for a complex world*. Boston, MA: Irwin/Mc Graw-Hill.
- Sterman, J. D. (1984). Appropriate summary statistics for evaluating the historical fit of system dynamics models. *Dynamica*, 10(2), 51–66.

- Theil, H. (1966). *Applied Economic forecasting*. Chicago: Rand McNally.
- Van Dyne, L., & Pierce, J. L. (2004). Psychological ownership and feelings of possession: Three field studies predicting employee attitudes and organizational citizenship behavior. *Journal of Organizational Behavior*, 25, 439–459.
- Vennix, J. A. M. (1996). *Group model building: Facilitating team learning using system dynamics*. Chichester: Wiley.
- Wolstenholme, E. F. (2003). Towards the definition and use of a core set of archetypal structures in system dynamics. *System Dynamics Review*, 19(1), 7–26.
- Zelman, W. N., Pink, G. H., & Matthias, C. B. (2003). Use of the balanced scorecard in health care. *Journal of Health Care Finance*, 29(4), 1–16.
- Ziegler, A., Lange, S., & Bender, R. (2007). Überlebenszeitanalyse: Der Log-Rang Test. *Deutsche Medizinische Wochenschrift*, 132, e39–e41.

Chapter 21

Public Policy Design for Climate Change Adaptation: A Dynamic Performance Management Approach to Enhance Resilience

Hugo Herrera

Abstract This chapter proposes dynamic performance management (DPM) as a suitable method to identify policies in the context of climate change adaptation. Namely, it focuses on the role it can play to support the analysis of how to enhance resilience of social and economic systems to climate change. While ‘resilience’ is a buzzword in the policymaking world, putting the concept into practice is still undeveloped. In a public administration focused on accountability, intangible outcomes of resilience represent a complication. The chapter discusses the findings and lessons from a case study applying the proposed approach. The results highlight the role of a dynamic performance approach to support stakeholder engagement, outcome-based policymaking and integrated solutions in the process of climate change adaptation.

Keywords Performance management · Resilience · System dynamics

21.1 Introduction

Effects of climate change are now hard to deny. In the past years, climate change has been manifested in a rise in temperatures and changes in rainfall seasonality around the globe. These effects of climate change have shocked our social and economic systems, exacerbating water scarcity, hunger and even social conflicts in many parts of the world. These events evidenced the dependence of social and economic systems on their natural counterparts and increased the interest in identifying ways to reduce vulnerabilities and foster successfully managed adaptation.

In this context, resilience has become a buzzword in the literature and politicians’ discourses (Adger 2000, 2009; Carpenter et al. 2001; Folke 2006). Resilience

H. Herrera (✉)

Department DEMS, University of Palermo, Bergen University, and Radboud University, Via Ugo Antonio Amico 3, 90100 Palermo, Italy
e-mail: hugojhdl@gmail.com

is the state of a system that withstands external changes due to its ability to absorb a certain amount of disturbances (Gallopín 2006; Gunderson 2000). In the social–ecological systems (SES) domain, resilience has been used to describe the properties of a system that allow it to continue providing desired outcomes, such as food, water supply and energy, even when the system has been affected by disturbances or shocks, like the effects of climate change.

The emergence of resilience has not gone unnoticed, capturing the interest of researchers and practitioners while remaining a cumbersome concept in the policymaking domain. In public administration, the concept of resilience still has a long way to go (Duit et al. 2010). Translating the ambition of making our social and economic system more resilient into effective policies presents a challenge to conventional policymaking and public managing approaches (Chapin III et al. 2009; Folke 2006).

Many critical voices have appeared in the contemporary environmental governance literature pointing out the current complications of transferring resilience thinking into practice. Three points of critique are as follows: (a) the lack of quantification and measures for resilience, (b) the alienation of resilience theories from the policymaking world and (c) simplified or rudimentary understanding of political processes (Duit et al. 2010).

This chapter focuses on the second point of critique, in particular in the context of the new public management (NPM) phenomenon and its output-based perspective. The literature on resilience is vast, with interesting and appealing concepts and theories, but so far it has failed to translate theories into real-world policies. The disconnection between resilience theory and resilience policymaking is manifested in (a) the simplified understanding of the political process described in the literature on resilience (Eriksen et al. 2015) and (b) its contradictions to policymaking and management processes in the public sector (Arnaboldi et al. 2015). The abstract and conceptually based approach of resilience particularly clashes with result-oriented views held within NPM. NPM is a development system that introduces practices used in the private sector into public administration (Arnaboldi et al. 2015). NPM is embedded in the public administration of many countries and government sectors by now, and it is and has been a key element supporting the implementation of output-oriented standards of performance (Arnaboldi et al. 2015; Pallot 1999).

In the subsequent sections, this chapter explores a dynamic performance approach (DPM) as a means to support the identification and design of policies for climate change adaptation based on the enhancement of resilience. Explicitly, it focuses on how DPM can be used as suitable bridge between the abstract concepts of resilience thinking and the concrete and measurable policies the public sector needs to design and account for. The resilience of food security to climate change in Huehuetenango, Guatemala, is used to illustrate the opportunities DPM might offer to policymakers analysing resilience within the public sector.

The chapter proceeds as follows: First, it describes the current position and challenges of resilience in the context of policymaking in the public sector. Next, it briefly describes how DPM can be used in the context of resilience analysis. The case study of food affordability in Huehuetenango, Guatemala, is used to illustrate

how DPM is applied in real-world contexts. Finally, the chapter discusses the opportunities to formalise resilience analysis in public administration using DPM.

21.2 Literature Review

21.2.1 *Resilience, Climate Change Adaptation and Public Policy*

Even if resilience is widely applied, a defining characteristic of the concept in SES literature is that ‘there is no single theoretical framework under which all resilience-related research is subordinated’ (Duit 2015, p. 5). Instead there is a diverse set of definitions, concepts and descriptions of what resilience means (Berkers et al. 2008; Chapin III et al. 2009; Folke et al. 2004; Walker et al. 2006, 2004); hence, scholars usually refer to the research related to resilience as resilience thinking rather than resilience theory (Walker and Salt 2012). In this chapter, resilience is understood, as defined by Walker and Meyers (2004), as the capacity of a system to absorb disturbance while retaining its essential function.

The recognition gained by resilience thinking in the context of climate change comes from the opportunities resilience could offer to the analysis. Resilience has become a common objective of climate change adaptation across a whole range of systems and activities, and it is an overarching concept in many strategies (Heller and Zavaleta 2009; Mawdsley et al. 2009).

In the public policy administration domain, the idea of resilience is not new. Already in the late 1980s, Wildavsky (1988) described resilience as a means to manage risk in modern societies; nowadays, it is a familiar concept in the crisis management literature (Aldrich 2012; Boin et al. 2010). However, the translation of resilience concepts into effective policies is still to a considerable extent unexplored in the public administration domain.

Current research on resilience policymaking is mainly found in the SES domain (Biggs et al. 2012; Chapin III et al. 2009). This literature focuses on the description of those social and natural properties of the system that are hypothesised to foster resilience, like redundancy, stakeholder participation and understanding of the system. The justification for these properties is found in case study research showing how the hypothesised properties enhanced the resilience of a particular outcome of the system to specific disturbances. Nevertheless, this justification is only at a conceptual level and rarely quantifies the impact of actions undertaken on the system, mainly because the way properties are enhanced is not clear. This is a downside in the current literature, since there is still a disconnection between the conceptual relations used to explain resilience and concrete policies, actions and plans to enhance it. The foregoing complicates the usage of resilience in the context of the NPM phenomenon focused on output-based performance.

21.2.2 New Public Management in the Public Sector and Output-Based Performance

The NPM phenomenon in the public sector has its inceptions in the late 1970s and early 1980s (Pallot 1999). NPM started in the United Kingdom and the municipal governments in the United States but rapidly expanded to other countries. However, it was only later, in the 1990s, that academics identified the common characteristics of these reforms and organised them under the label of ‘new public management’ (Dunsire 1995). NPM assumes that the management tasks in the public sector are not significantly different from the managerial tasks in private sector organisations and therefore that private sector techniques can be usefully applied in public administration (Pallot 1999). The introduction of private sector techniques in public administration was justified by the aim of enhancing public sector flexibility, accountability and control; a client and service orientation; a strengthened capacity for developing strategy and policy; introducing competition and other market elements; and changed relationships with other levels of government (Lynn 1998).

NPM was grounded, to a large extent, in the hypothesis that the poor performance of bureaucratic structures could be improved if the public sector would act more like its private counterpart. For instance, performance would improve if the public sector would be more product- instead of function-oriented or if management objectives would become dominant over legal arrangements. Intentionally, the NPM phenomenon resulted in a shift from a process-oriented perspective to an output-based one focused on results, efficiencies and the value for money (Pallot 1999; Vries and Nemeč 2013).

Nevertheless, in practice, the public sector faces, by far, more difficult problems than any business in the private sector (Pallot 1999, p. 22). Many of these problems arise from complex systems, characterised by an underlying causal structure of accumulation processes, the core of any dynamic system, that are interrelated by way of nonlinear feedback structures that cut across sectors and disciplines. The resulting various delays are spread across a system, making the root cause of a problem nearly inaccessible and preventing the identification of the timing and dosage of effective interventions. In addition, circular causality, that is, feedback, leads us into circular arguments, made meaningful only when we recognise the associated delays. Such is the case of building resilience for climate change effects, where the results of the policies implemented are only observable, if so, after long time periods of time and depend on the management of the feedback loop structures influencing the system.

Recognising the complexity of public sector problems is needed in order to understand and effectively act on the dynamics of the SES affected by climate change. Rather than assuming a social system characterised by stability and equilibrium, the analytical focus is placed on understanding processes of change and surprises and on how governance arrangements try to cope with and adapt to a constantly dynamic and evolving environment (Duit et al. 2010).

NPM approaches and their focus on results often fail to capture the dynamic complexity of managerial decisions by underestimating a number of relevant factors

influencing policy performance (Arnaboldi et al. 2015; Bianchi 2010). In fact, by narrowing the measures of policy performance to only the outputs of the system, NPM might constrain the implementation of policies to enhance system resilience, as they do not deliver tangible results measurable in the short term. An additional complication results from the fact that, as mentioned before, the resilience literature is not clear on how to measure resilience or what to measure. This lack of quantification and operationalisation of the concept of resilience is a significant challenge for NPM, because the benefits of policies are hard to assess and there is not a clear framework in the literature to do so.

To summarise, so far, the concept of resilience has been mainly developed in the SES as a metaphor of how systems should behave (Duit 2015). The idea of resilience in public administration is still awkward mainly due to the discrepancies between NPM perspectives and the inherent complexity of resilience. If resilience is to be used widely in public administration, policymakers need to be able to connect policies with their impact on resilience, compare their effectiveness (e.g. in term of added value for money) and measure their performance. If any progress toward a more active incorporation of resilience perspectives into the public policies were to occur, a new approach for performance management would be required.

21.3 Methodology: Dynamic Performance Management Approach

This chapter explores DPM as approach to bridge the literature on resilience thinking and the public sector policymaking world. DPM is a combination of performance management approaches and system dynamics (SD) (Bianchi 2016). DPM supports policymaking process by modelling organisational systems (in SD models) and using simulation techniques to understand the behaviour of the complex systems public policies deal with (Bianchi and Rivenbark 2012; Bianchi and Tomaselli 2015). The contribution of DPM is to help policymakers assess the middle and long-term impacts of their actions in the system outputs by placing the measure of performance in a broader context of the system (Bianchi and Tomaselli 2015).

In the analysis of resilience, DPM provides a means for discussing the concept of resilience in a more operational manner. Figure 21.1 shows the three interconnected views of system performance covered by DPM (Bianchi 2016) and the way they merge with the analysis of resilience. The objective view includes activities and processes influencing the system behaviour, performed by different stakeholders within and outside public administration. The subjective view includes performance goals, performance measures and key indicators defined by higher governmental policies and strategies. While these two views are rather common in traditional performance management, the third one, the instrumental view, is an important addition of the DPM approach. The instrumental view explicitly represents activities, processes, products and their relationships in terms of strategic resources and performance drivers. Performance drivers are the mechanisms conditioning the

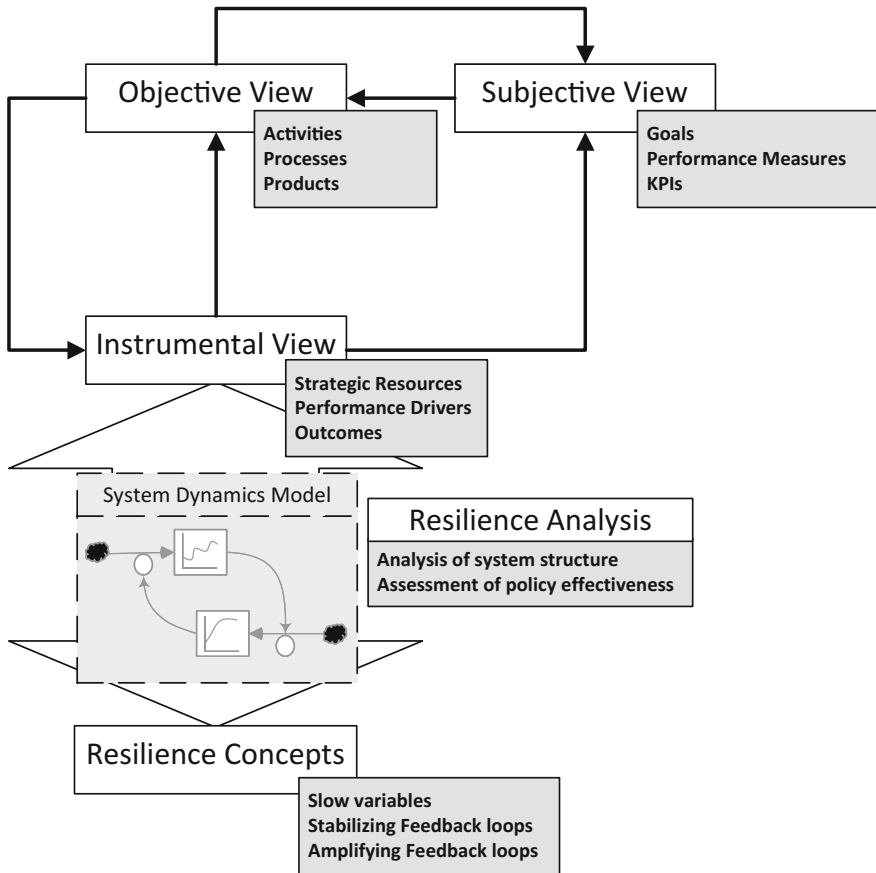


Fig. 21.1 Dynamic performance management framework in the context of resilience analysis

system outcomes and outputs, while strategic resources are the means supporting the performance drivers. By identifying what the links between the system goals in terms of outcomes and the performance drivers enabling them are, policymakers can identify important performance measures and more effective policies to improve the system responses. The foregoing, represented in an SD model, allows simulation of the behaviour of key elements in the system, to anticipate pitfalls of the policies proposed and to identify opportunities.

Moreover, the instrumental view is also a very suitable tool to represent key concepts of resilience, qualitative and formal simulation models. Systemic effects described in the resilience literature are easily grasped from the instrumental view. These concepts can be linked back to strategic resources, performance drivers and concrete processes. Then, it is possible to draw concrete action plans, identify key performance indicators and measure performance of policies to enhance resilience. DPM acts here as a bridge between accountable elements needed to manage policies in public administration and abstract concepts needed to describe and interpret the resilience of SES.

21.3.1 Resilience Analysis

The SD model, built within a DPM framework, can be used in the analysis of resilience to (a) analyse the system structure, looking for the mechanisms driving the system resilience and (b) explore in a systematic and quantifiable way the effectiveness of different policies. The analysis of the system structure focuses on identifying slow variables (Chapin III et al. 2009). Slow variables are variables that strongly influence the system but remain relatively constant over time (Chapin III et al. 2009).

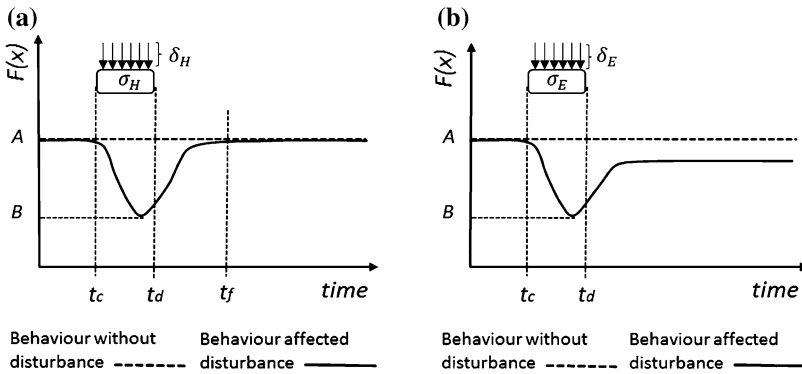
While many other analytical frameworks for assessment of resilience require making substantial abstractions, simulation can offer a more practical one (Schattka et al. 2016). Simulations can be used to assess the effect of policies on the resilience of the system by reproducing the behaviour of the system outcomes while affected by a given disturbance. The system outcomes can be represented by a quantifiable and time-dependent outcome function $F(x)$ from which distinctive properties can be measured (Barker et al. 2013; Henry and Emmanuel Ramirez-Marquez 2012). In this chapter one measures the five characteristics described by Herrera and Kopainsky (2015). These characteristics offer a comprehensive understanding of system resilience and are easily elicited from the simulated behaviour. Table 21.1 presents the resilient characteristics measured in this chapter to operationalise resilience.

Main parameters needed to calculate the resilience measures described in Table 21.1 are as follows:

- δ_H Amount of disturbance necessary to alter the behaviour of the system
- δ_E Amount of disturbance necessary to move the system to a different equilibrium
- t_d Time when the disturbance starts to affect the system
- t_c Time when the disturbance stops
- t_f Time when the system fully recovers.

Table 21.1 Characteristics of resilience and how to measure them

Measure	Description	Equation
Hardness (σ_H)	The ability of the system to withstand a disturbance σ without presenting a change in the performance of the outcome function $F(x)$.	$(\sigma_H) = \delta_H \times (t_d - t_c)$
Recover rapidity (\bar{R})	The average rate at which a system returns to equilibrium after a disturbance σ (Martin et al. 2011; Pimm 1984)	$(\bar{R}) = \frac{A-B}{t_f-t_d}$
Robustness ($\bar{\rho}$)	The system's ability to withstand big disturbances σ without significant loss of performance (Attoh-Okine et al. 2009)	$(\bar{\rho}) = \frac{\sigma}{A-B}$
Elasticity (σ_E)	The ability of the system to withstand a disturbance σ without changing to a different steady state (Holling 1996; Holling, Gunderson, & Peterson 2002)	$(\sigma_E) = \delta_E \times (t_d - t_c)$
Index of resilience (I_R)	The probability of keeping the current steady state or regime (Holling, 1996; Holling, Gunderson, & Peterson 2002; Martin et al. 2011).	$(I_R) = P(\sigma \leq \sigma_E)$



A: value of the outcome function $F(x)$ for the baseline system
 B: lower value of the outcome function $F(x)$ when the system is affected by a disturbance

Fig. 21.2 Two hypothetical pairs of responses to a disturbance and the parameters needed to calculate five characteristics of resilience. **a** System affected by a disturbance σ_H big enough to change the performance of the outcome function $F(x)$. **b** System affected by a disturbance σ_E big enough to change the behaviour of outcome function $F(x)$ to a different steady state

These parameters can be measured from the simulated behaviour of the outcome function $F(x)$ as it is illustrated in Figs. 21.2a and 21.2b. Since the σ_H and σ_E represent system thresholds—points or levels at which a significant variation in the behaviour is manifested—it is necessary to simulate the system response to a wide range of disturbances in order to identify them. The first threshold is the smaller σ that produces a change in the behaviour (hardness σ_H), and the next one is the smaller σ that results in a new equilibrium for the system (elasticity σ_E). The behaviour produced by disturbances larger than σ_H but smaller than σ_E are used to calculate the average recovery rapidity and robustness. Finally, the index of resilience (I_R) is the probability that σ would be larger than σ_E , and it can be easily calculated using the probability distribution function of σ .

21.4 Case Study: Policies to Enhance Resilience of Food Affordability in Huehuetenango, Guatemala

The Inter-American Development Bank has identified Guatemala among the top ten countries most vulnerable to climate change (World Bank 2003). Guatemala has been severely affected by climate change, mainly experiencing a drastic change in average rainfalls, which have caused both droughts and floods in magnitudes that have not been seen before. Relying on agriculture as its primary economic activity and 26% of its GDP, Guatemala has a vulnerability to climate change that is a high risk to its economy. Additionally, Guatemala is the fourth most susceptible nation

to natural disasters and suffers the eight highest incidences of childhood malnutrition in the world, according to UNICEF (2015). Guatemala's chronic malnutrition, an accepted measure of food insecurity, is the third worst in the (World Bank 2003). This combination of factors places the country's food security at high risk.

Huehuetenango has an area of 7400 km² (INE 2013b) and is located in the northwest region of Guatemala, on the border with southern Mexico. Huehuetenango is one of the poorest and most vulnerable districts of Guatemala, with a population estimated at 1.1 million inhabitants in 2014, 67.6% of them under the poverty line (INE 2013a). Huehuetenango's main economic activities are the mining of silver and gold and the production of coffee. The production of maize is, nevertheless, an important activity for self-consumption. The majority of the population is indigenous, of the Mam and Quechi ethnicities, and has a cultural dependence on maize as a main source of calories (71.2% of their basic grains consumption), especially for those in the rural areas, representing around 52% of the total population (Camposeco et al. 2008). Moreover, farming techniques are rudimentary, based on knowledge that has been passed down from previous generations and on the use of simple tools and principles. The effect of these poor conditions and basic techniques became evident in the lowest average yield historically recorded on small farms (between 1.5 and 2.3 tonnes/km² year). Since yields are low, maize produced is mainly used for self-consumption or local trading, alleviating local food vulnerability. Households' weak purchasing power and poor access by road make the local market less attractive for foreign producers and highly dependent of the local production.

21.4.1 *Subjective View*

The policymakers' objective was to identify policies to enhance the resilience of food affordability in maize-based systems to the increasing variation in the rainfall in the district of Huehuetenango. Henceforth, a maize-based system is understood as the system formed by (a) the small farmer producers of maize in the region, (b) the households in poverty who mainly produce and consume that local maize, (c) the local maize supply chain and (d) the ecosystem (soil and water) in which the maize is produced.

The policymakers' objective defines the scope of the analysis by identifying the variables to analyse (a) the outcome function $F(x)$ is food affordability and (b) the disturbance (σ) is the reduction in rainfall. Affordability was measured in the model as an index that reflects the ratio between the theoretical amount of maize required by a household and the actual amount the household can purchase. The units of this index are dimensionless but will be addressed in this chapter as points on the affordability index to avoid confusion. If the affordability is 1.00, it means that, on average, households can buy all the maize they need to cover their daily requirements. Alternatively, 0.50 point on the affordability index means that, on average,

households can only buy half of the total maize they need to cover their daily requirements. The reduction of rainfall indirectly causes starvation of many families by destroying the crops of small-scale farmers. There is an estimate that rainfall has decreased 31% in the last 15 years, and projections expect that 2016 will be the driest year ever in Guatemala (Gándara 2016).

21.4.2 Objective View

Four stakeholders were identified as the main groups intervening in the maize system of Huehuetenango, Guatemala. In addition to the government and the farmers, the supply chain (big producers and retailers of maize) and NGOs supporting farmers with technical and financial assistance are also important players. The areas of influence, relevant processes and products for each stakeholder group are presented in Table 21.2.

21.4.3 Instrumental View

The process by which the different stakeholders described in Table 21.2 interact with the environment is mapped on a diagram using an instrumental view. The

Table 21.2 Main stakeholders' areas of influence, activities and products

Stakeholder	Areas of Influence	Activities and processes	Products
Ministry of Agriculture	Public policies and public resources	<ul style="list-style-type: none"> • Policy design and implementation • Technical training for farmers • Agriculture market regulation 	Technical and economic framework for maize production (recommended seeds, practices, subsidies, technical assistance)
NGO	Hard and soft resources for supporting farmers	<ul style="list-style-type: none"> • Technical training for farmers • Donations and aids for farmers 	Capacitation programmers, donations of seeds and other inputs
Households farmers	Household economic assets and decisions	<ul style="list-style-type: none"> • Maize production • Maize trade • Livestock farming 	Allocation of resources (household cash, time, efforts and land)
Supply chain (big producers and retailers of maize)	Market	<ul style="list-style-type: none"> • Maize production • Maize consumption 	Market conditions (price, supply, quality)

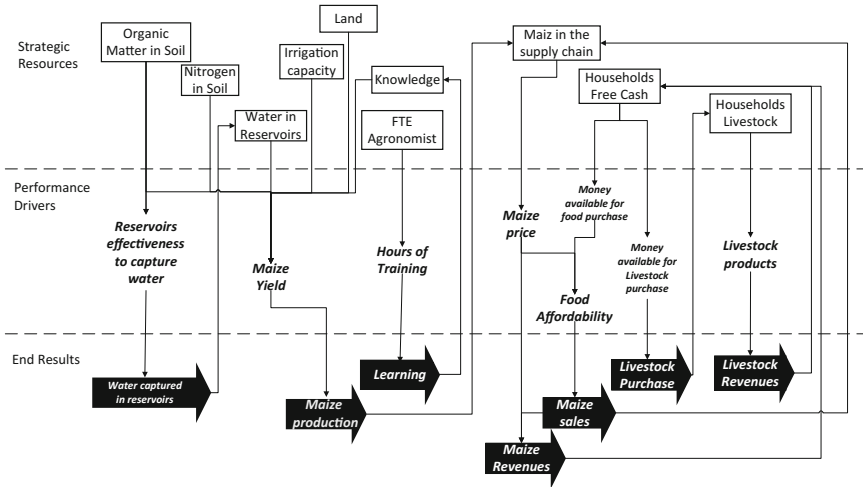


Fig. 21.3 A summarised three-view framework for a dynamic performance management approach to the maize system of Huehuetenango. *Note* 1 FTE (full-time equivalent) = 42 hours per week

purpose of this mapping activity is to make these interactions explicit and to identify the key performance drivers. Key performance drivers are supported by strategic resources. Strategic resources can be part of the ecological dimension (e.g. nitrogen in soil), economic dimension (e.g. households’ free cash) or social dimension (e.g. farmers’ knowledge) of the maize system. The diagram summarising the system interactions analysed in the instrumental view is presented in Fig. 21.3. Note the relationships in the diagram: strategic resources support performance drivers, performance drivers produce end results and end results (in most cases) build strategic resources in a virtuous circle.

21.4.4 System Dynamics Model

The relationships captured in Fig. 21.3 were used to produce an SD simulation model. The structure of the model is represented in a stock-and-flow diagram in Fig. 21.4. The strategic resources identified before are represented in the SD model as ‘stocks’ (the rectangular boxes). Stocks are variables that represent accumulations (Cote and Nightingale 2012). The ‘households’ free cash’, for instance, is an economic strategic resource affected by households’ ‘revenues’ and ‘expenditures’. They are represented by double arrows coming in and out of the stock. The double arrows represent the variables increasing or depleting the strategic resources. These variables are the rate at which the strategic resource grows and decreases. The households’ free cash defines the ‘money available for purchasing food’

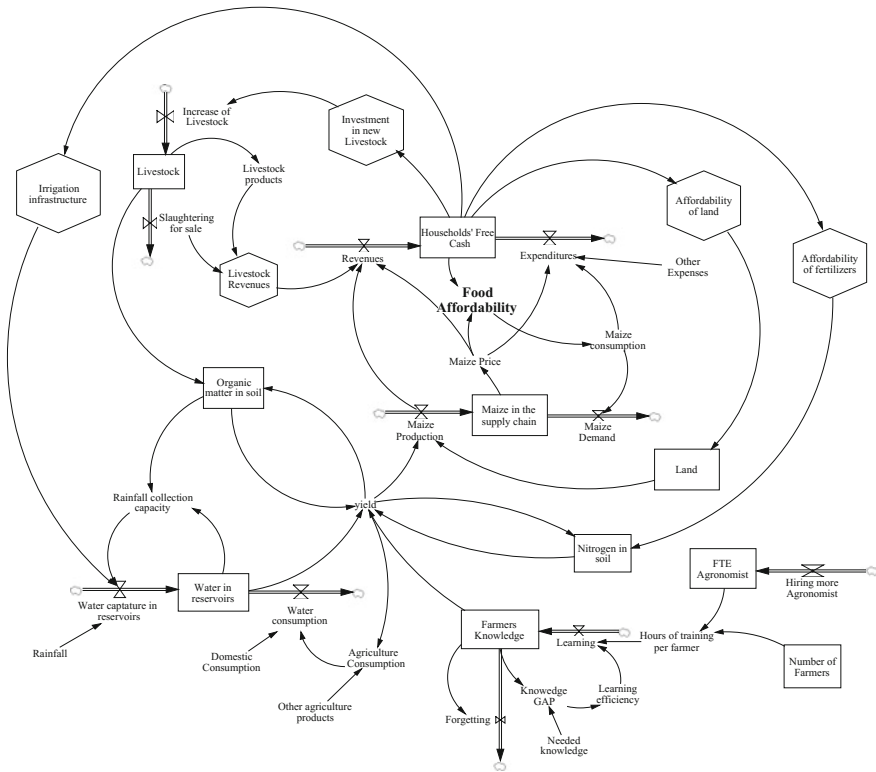


Fig. 21.4 Simplified stock-and-flow diagram representing the main structure of the system dynamics model build for the case study. *Note* Hexagons represent structures included in the model but not presented in the figure. *FTE* full-time equivalent = 42 hours per week

(the performance driver in Fig. 21.3). Food affordability, the main goal of the system, is the result of the relation between the price of maize and the money available for purchasing food (Fig. 21.4). Food affordability is not only a main goal of the system but also a performance driver by itself, influencing end results like the ‘maize demand’ and the household’s ‘expenditures’.

21.4.5 Resilience Analysis

The resilience analysis focuses on identifying slow variables and feedback loops that have an effect on the resilience of food affordability to reductions in rainfall. It is important to note that even the final goal of the policies is to increase the resilience of food affordability, this cannot increase by simply acting on this variable, but rather by enhancing the strategic resources influencing the performance drivers that define the response of food affordability to a decrease in rainfall.

Our analysis focused on three slow variables as points of intervention, ‘nitrogen in the soil’, ‘livestock’ and ‘water in reservoirs’. The nitrogen coming from organic and artificial fertilisers accumulates in the soil and is depleted by the crops absorbing it and using it to produce standing biomass (see Fig. 21.5). The crop’s yield depends on a large extent on the amount of nitrogen in the soil in a nonlinear relationship. Nitrogen in the soil can be increased by adding fertilisers to the soil or some forms of organic matter (e.g. crop residues from leguminous plants). Namely, subsidies for fertilisers were proposed as policy to increase the nitrogen in the soil.

In Huehuetenango’s maize system, livestock is an important resource to generate alternative sources of revenues for households depending on maize production. Livestock can generate revenues through slaughtering then selling the meat and/or by selling products produced by the livestock (e.g. eggs, milk, etc.). Figure 21.6 illustrates how ‘livestock revenues’, ‘households’ free cash’ and ‘livestock’ are connected on a virtuous cycle. In the context of resilience, the livestock acts as a buffer in case a disturbance affects the maize production, by providing an alternative source of revenues not directly affected by that disturbance.

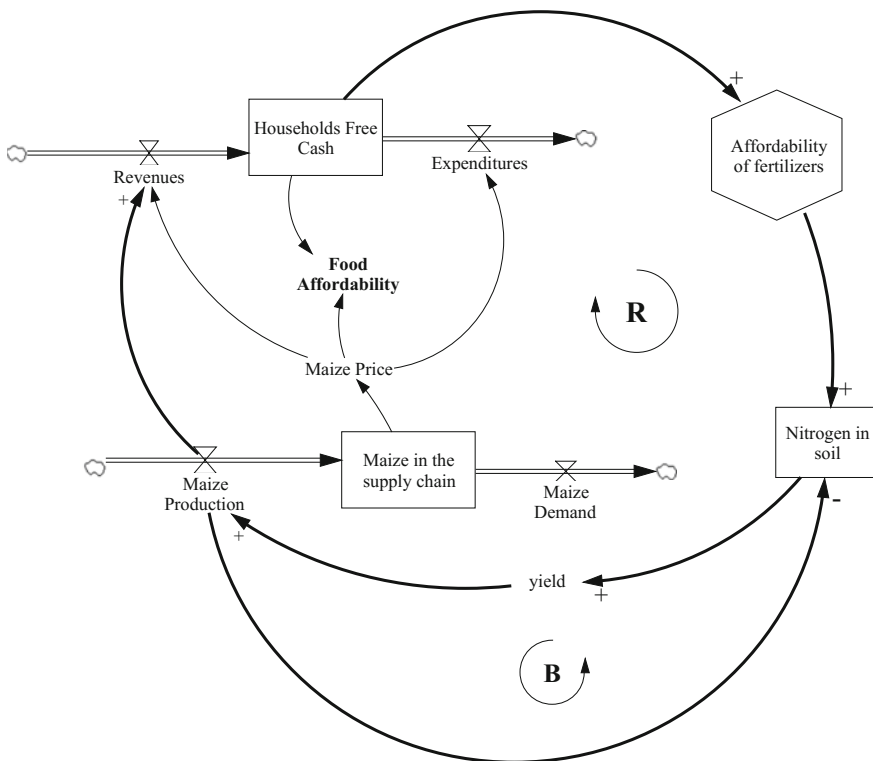


Fig. 21.5 Stock-and-flow diagram representing the subsection of the model related to nitrogen in the soil. *Note* For presentation purposes, some elements of the structure included in the model are not presented in the diagram

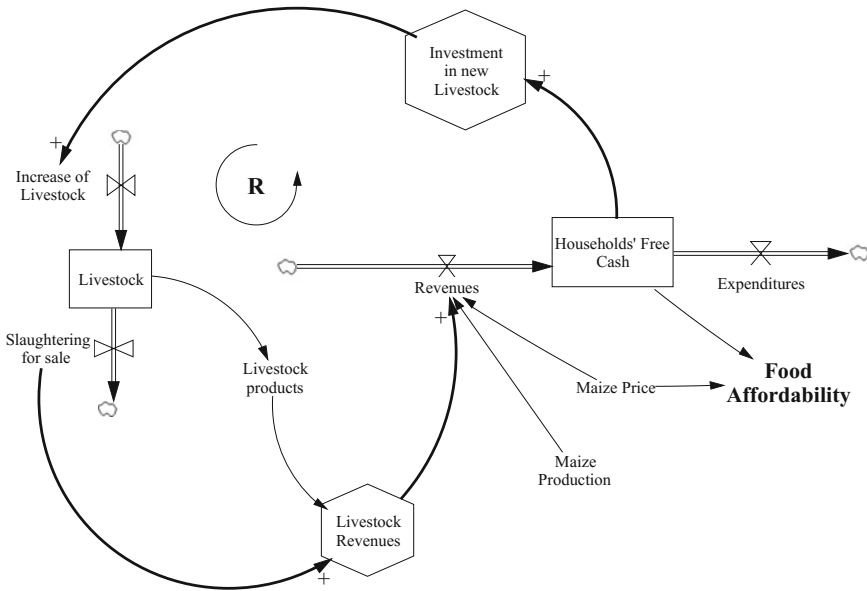


Fig. 21.6 Stock-and-flow diagram representing the performance drivers of the households' free cash. Note: For presentation purposes, some elements of the structure included in the model are not presented in the diagram

Policy interventions can provide farmers with incentives to increase the investment in additional livestock. For instance, government can (a) provide subsidies for livestock purchases, (b) donate livestock to the household and (c) provide technical support to improve the management of the current livestock.

It is worth mentioning that livestock, especially cattle, is one of the biggest sources of greenhouse gases (namely methane). Greenhouse gases are indirectly responsible for climate change effects. Policies increasing livestock might have counterintuitive consequences, diminishing resilience in the long term. To prevent these unintended results, options like small-scale organic poultry (with low methane emissions) should be prioritised over cattle.

Finally, and closely related to the disturbance, is the water in reservoirs. The water in natural basins (e.g. rivers, lakes and the soil itself) and artificial reservoirs (e.g. tanks) is a key resource for maize production (see Fig. 21.7). Maize, being one of the most water-demanding crops, requires an appropriate and constant supply of it. The water reservoirs depend on (a) the capitation rate (the amount of water captured) and (b) the consumption rate. The capitation rate depends on the amount of rain but also the capitation capacity of the reservoir itself. Eroded soils, for instance, are ineffective at capturing the water that rains on them. In this case, the main uses of water are agriculture and domestic consumption. Considering the timescales modelled (10 years) and with the purpose of simplifying the analysis, domestic consumption is here considered constant. Alternative policies to affect the

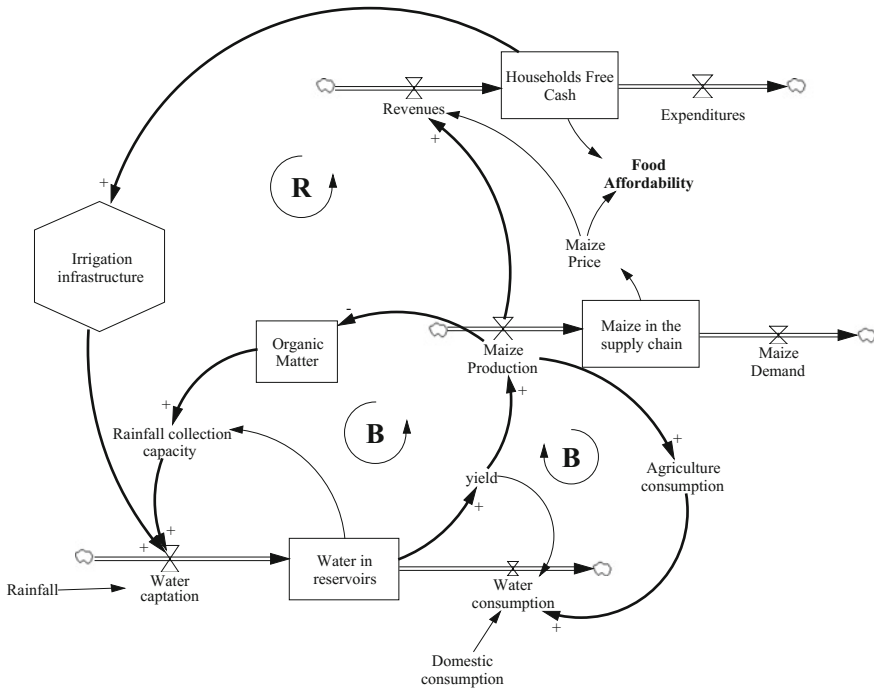


Fig. 21.7 Stock-and-flow diagram representing the performance drivers of maize production

water in reservoirs are (a) to build water storage capacity to be able to capture more water when the rainfall is abundant, (b) to encourage the utilisation of maize varieties that require less water and (c) to increase the soil’s capacity to retain rainfall by adding organic matter to it.

In summary, three policies were identified (a) to increase the nitrogen in the soil by encouraging the use of fertilisers (Policy 1), (b) to increase revenues livestock revenues by offering subsidies and training to farmers (Policy 2) and (c) to increase water storage capacity (e.g. cisterns) by offering financial and technical aid to vulnerable farmers (Policy 3). These three policies are simulated to assess their benefits and effectiveness. The current system and the systems including the policies proposed were simulated considering a rainfall reduction with a magnitude between 0 and 35% over a period of 5 years. The simulated behaviour is used to measure the impact of the proposed policies on the resilience as described in Sect. 21.3. The calculated values for the baseline scenario (scenario with no policy) and the proposed policies are presented in Table 21.3.

Different policy recommendations can be made when looking at the different performance scores in Table 21.3. For instance, if the aim is to enable farmers to recover quickly after a drought period, Policy 1 (to increase the nitrogen in the soil) seems more appropriate because has the higher ‘recover rapidity’. Alternatively, Policy 2 (increase livestock revenues) is preferable if the aim is to have a flexible

Table 21.3 Measures of the resilience of food affordability of the maize system in Huehuetenango

Measure		Baseline	Policy 1	Policy 2	Policy 3
Hardness	(% Annual rainfall variation)	12%	18%	19%	21%
Recover rapidity	(Points of affordability index/year)	4.12	6.71	6.65	5.23
Robustness	(% Annual rainfall variation/points of affordability index)	0.61	0.88	0.98	0.85
Elasticity	(% annual rainfall variation)	34%	43%	73%	41%
Resilience index	(% Probability to maintain regime)	73%	78%	95%	68%

Policy 1: to increase the nitrogen in the soil, Policy 2: to increase livestock revenues, Policy 3: to increase water storage capacity (e.g. cisterns)

Numbers in bold are the highest value for each measure

system, able to withstand extreme drought without compromising future subsistence of the farmers. Policy 2 is the one with the higher scores for ‘elasticity’ and ‘index of resilience’, both indirect measures of the system flexibility. The ambiguity of resilience requires a dialogue about the explicit goals different stakeholders have for the system and their understanding of resilience. DPM might facilitate this dialogue by offering an operational and quantifiable framework.

As good as the analysis might be, it might yield practical benefits only when the proposed policies are transformed into concrete plans. Concrete plans for the policy implementation include activities, timetables and resources needed in the process. These activities are part of the objective view summarised in the DPM diagram shown in Fig. 21.8. The understanding of the specific activities needed to implement each policy allows for an estimate of how much time, money and resources in general are needed for their implementation (see Fig. 21.8). The original subjective view, including the overall goal of the system, was complemented by adding specific key performance indicators related to the important end results of each policy.

Policy recommendations can also be assessed by evaluating the benefits of each policy against its costs. The high-level plans shown in Fig. 21.8 were used to estimate the costs and the net present value (NPV) of each policy. The NPV is used to account for the cost of different policies in comparable settings. NPV is also used to calculate the value-for-money ratio for each of the policies. Value for money is used as a measure of the benefits delivered by each policy versus its cost in NPV terms. The benefits delivered by each policy are the positive changes observed in the response of the outcome function to the disturbance affecting the system. These changes can be deduced from the differences in the values of each characteristic measured with the policy in place and without it.

Table 21.4 shows the NPV cost for each one of the three policies analysed and their value-for-money ratio (expressed in benefits per million USD). The results indicate that Policy 2 (increasing livestock reserves) is the one providing higher value for money, because it delivers more benefits per each USD invested.

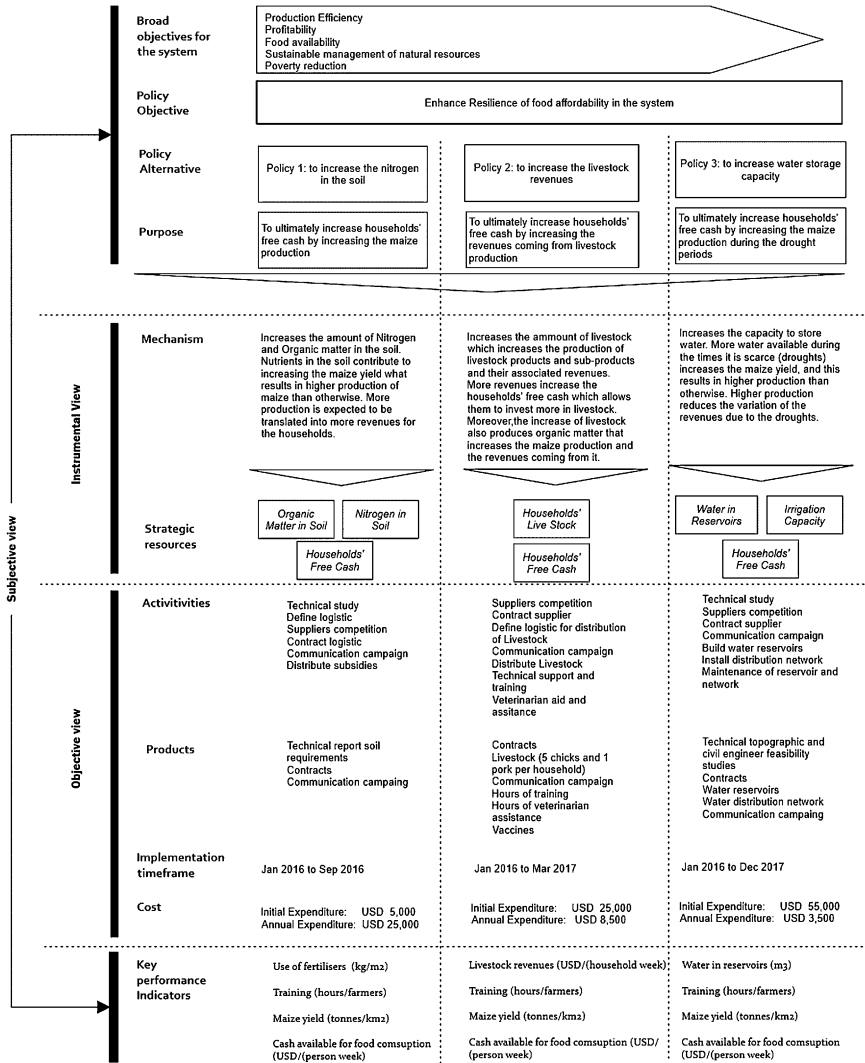


Fig. 21.8 A summarised three-view description of the policies to enhance food affordability resilience identified previously in this chapter

21.5 Discussion

While resilience is trending in many academic fields, this chapter focuses on the practical applications of the concept of resilience in public policy design. The case study described above, even if only an example, shows that there are opportunities for bringing together in one single approach the abstraction needed to analyse resilience and the concreteness needed to implement it.

Table 21.4 Net present values and value-for-money ratio for the policies proposed to enhance the resilience of food affordability in the maize system in Huehuetenango

	NPV (USD)	Value for money (benefits per millions of USD)				
		Hardness	Recover rapidity	Robustness	Elasticity	Resilience index
Policy 1	65,000	0.92	39.85	4.15	1.38	0.77
Policy 2	71,000	0.99	35.63	5.21	5.49	3.10
Policy 3	95,000	0.95	11.68	2.53	0.74	<i>(-0.53)</i>

Value for money is calculated as the difference in the measure between the policy and the baseline divided by the NPV expended to achieve the difference. NPV calculated using a discount rate of 3.5% per year. Policy 1: to increase the nitrogen in the soil, Policy 2: to increase livestock revenues, Policy 3: to increase water storage capacity (e.g. cisterns)

Numbers in bold are the highest value for each measure. Negative values in italics

The design and implementation of policies that aim to enhance resilience requires transparent means to connect actions and effects. In a DPM approach, the link between concrete plans and the effects of resilience is transparent, because slow variables, strategic resources, performance drivers and concrete processes harmonically coexist. For instance, the slow variables identified in the analysis of resilience (nitrogen in soil, livestock and water in reservoirs) correspond to strategic resources supporting processes driving the system's behaviour. Experts and stakeholders can smoothly navigate between concepts through different levels of abstraction. The process moves easily from policy design to implementation to measure of performance, all without risks of losing ownership or accountability in the process.

In this way, the usage of DPM uncovers opportunities to formalise resilience analysis in public administration. DPM acts as transitional tool facilitating dialogue and encouraging policymakers to (a) define resilience in terms of objective and measurable targets, (b) describe policies with regard to intermediate products and services related to concrete activities and processes and (c) analyse the system in terms of strategic resources and performance drivers.

21.6 Conclusions and Further Research

There is potential for resilience to contribute to public policymaking around climate change adaptation. However, the resilience literature so far is too abstract, and the policy recommendations are unfamiliar to the policymaking process and awkward in their implementation. There is still substantial work to do in order to effectively integrate resilience thinking into public administration.

DPM is a promising approach for bridging abstract resilience concepts with the policymaking world and public administration. The instrumental view of the DPM approach connects concrete activities and processes with abstract concepts of resilience, namely, DPM connects through different levels of analysis, activities, performance drivers, strategic resources and slow variables, allowing navigation

back and forth between the concrete policies and abstract mechanisms to enhance resilience. This transparent link between resilience and public policy domains can boost resilience as a sound framework for policymaking and climate change adaptation.

The case study described in this chapter illustrates the potential of DPM and shows how this approach can support a more robust resilience analysis able to withstand the criticism of those in control of public funds. Nevertheless, this case is the only one example of the opportunities for the use of DPM, and more research is needed. Further work should include a wider range of applications, different types of stakeholder engagements and follow-up to policies implemented.

References

- Adger, W. N. (2000). Social and ecological resilience: Are they related? *Progress in Human Geography*, 24(3), 347–364.
- Adger, W. N. (2009). Social Capital, Collective Action, and Adaptation to Climate Change. *Economic Geography*, 79(4), 387–404.
- Aldrich, D. P. (2012). *Building resilience: Social capital in post-disaster recovery*. Chicago: University of Chicago Press.
- Arnaboldi, M., Lapsley, I., & Steccolini, I. (2015). Performance management in the public sector: The ultimate challenge. *International Journal of Public Sector Management*, 31(1), 1–22.
- Attoh-Okine, N. O., Cooper, A. T., & Mensah, S. (2009). Formulation of resilience index of urban infrastructure using belief functions. *IEEE Systems Journal*, 3(2), 147–153.
- Barker, K., Ramirez-Marquez, J. E., & Rocco, C. M. (2013). Resilience-based network component importance measures. *Reliability Engineering and System Safety*, 117, 89–97.
- Berkers, F., Colding, J., & Folke, C. (2008). *Navigating social-ecological systems: Building resilience for complexity and change*. Cambridge: Cambridge University Press.
- Bianchi, C. (2010). Improving performance and fostering accountability in the public sector through system dynamics modelling: From an “external” to an “internal” perspective. *Systems Research and Behavioral Science*, 27(4), 361–384.
- Bianchi, C. (2016). *Dynamic performance management*. Berlin: Springer.
- Bianchi, C., & Rivenbark, W. C. (2012). A comparative analysis of performance management systems: The cases of sicily and North Carolina. *Public Performance and Management Review*, 35(3), 509–526.
- Bianchi, C., & Tomaselli, S. (2015). A dynamic performance management approach to support local strategic planning. *International Review of Public Administration*, 20(4), 370–385.
- Biggs, R., Schluter, M., Biggs, D., Bohensky, E. L., BurnSilver, S., Cundill, G., et al. (2012). Toward principles for enhancing the resilience of ecosystem services. *Annual Review of Environmental Resources*, 37, 421–448.
- Boin, A., Comfort, K. L., & Demchak, C. C. (2010). *The Rise of resilience. In designing resilience: Preparing for extreme events* (Vol. 1–13). Pittsburgh, PA: Pittsburgh University Press.
- Camposco, M., Thomas, M., & Kreyndar, W. (2008). *Huehuetenango en Cifras*. Huehuetenango: Centro de Estudios y Documentacion de la Frontera Occidental Guatemala, p 13–31.
- Carpenter, S., Walker, B., Anderies, M. J., & Abel, N. (2001). From metaphor to measurement: Resilience of what to what? *Ecosystems*, 4(8), 765–781.

- Chapin III F. S., Kofinas G. P., Folke C. (2009). *Principles of Ecosystem Stewardship: Resilience-Based Natural Resource Management in a Changing World*. New York: Springer Science & Business Media.
- Cote, M., & Nightingale, A. J. (2012). Resilience thinking meets social theory Situating social change in socio-ecological systems (SES) research. *Progress in Human Geography*, 36(4), 475–489.
- Duit A. (2015). Resilience Thinking: Lessons for Public Administration. *Public Administration*, 94 (2), 364–380.
- Duit, A., Galaz, V., Eckerberg, K., & Ebbesson, J. (2010). Governance, complexity, and resilience. *Global Environmental Change*, 20(3), 363–368.
- Dunsire, A. (1995). Administrative theory in the 1980s: A viewpoint. *Public Administration*, 73 (1), 17–40.
- Eriksen, S. H., Nightingale, A. J., & Eakin, H. (2015). Reframing adaptation: The political nature of climate change adaptation. *Global Environmental Change*, 35, 523–533.
- Folke, C. (2006). Resilience: The emergence of a perspective for social-ecological systems and analyses. *Global Environmental Change*, 16(3), 253–267.
- Folke, C., Carpenter, S., Walker, B., Scheffer, M., & Elmqvist, T. (2004). Regime shifts, resilience, and biodiversity in ecosystem management. *Annual Review of Ecology Evolution and Systematics*, 35, 557–581.
- Gallopín, G. C. (2006). Linkages between vulnerability, resilience, and adaptive capacity. *Global Environmental Change*, 16(3), 293–303.
- Gándara, P. (2016). *El Niño y La Niña provocarán sequía y lluvia torrencial*. Retrieved from <http://www.prensalibre.com/economia/fenomenos-el-nio-y-la-nia-afectan-en-2016>. Accessed 15 May 2016.
- Gunderson, L. H. (2000). Ecological resilience in theory and application. *Annual Review of Ecology and Systematics*, 31, 425–439.
- Heller, N. E., & Zavaleta, E. S. (2009). Biodiversity management in the face of climate change: A review of 22 years of recommendations. *Biological Conservation*, 142(1), 14–32.
- Henry, D., & Emmanuel, Ramirez-Marquez J. (2012). Generic metrics and quantitative approaches for system resilience as a function of time. *Reliability Engineering and System Safety*, 99, 114–122.
- Herrera, H., & Kopainsky, B. (2015). Rethinking agriculture in a shrinking world: Operationalization of resilience with a System Dynamics perspective. In *33rd International Conference of the System Dynamics Society 2015 Volume 2*. Albany: System Dynamics Society, p 1343–1360.
- Holling, C. S. (1996). Engineering resilience versus ecological resilience. In Schulze P. (Ed.). *Engineering Within Ecological Constraints*. Washington DC: National Academy Press, p 31–44.
- Holling, C. S., Gunderson, L., & Peterson, G. (2002). Sustainability and panarchies. In L. H. Gunderson & C. S. Holling (Eds.), *Panarchy: Understanding transformations in human and natural systems*. Washington DC: Island Press, p 63–102.
- INE. (2013a). *Mapas de pobreza Rural en Guatemala 2011*. Guatemala: Instituto Nacional de Estadística. Retrieved from <http://www.ine.gob.gt/sistema/uploads/2014/01/10/ifRRpEnf0cjUfRZGhyXD7RQj77EQH2Er.pdf>.
- INE. (2013b). *Caracterización estadística República de Guatemala 2012*. Guatemala: Instituto Nacional de Estadística. Retrieved from <http://www.ine.gob.gt/sistema/uploads/2014/02/26/5eTCcFIHERnaNveUmm3iabXHakgXtw0C.pdf>.
- Lynn, L. Jr. (1998). The new public management: How to transform a theme into a legacy. *Public Administration Review*, 58(3), 231–237.
- Martin, S., Deffuant, G., & Calabrese, J. M. (2011). Defining resilience mathematically: From attractors to viability. *Viability and Resilience of Complex Systems* (pp. 15–36). Berlin: Springer.

- Mawdsley, J. R., O'Malley, R., & Ojima, D. S. (2009). A review of climate-change adaptation strategies for wildlife management and biodiversity conservation. *Conservation Biology*, 23(5), 1080–1089.
- Pallot, J. (1999). Beyond NPM: Developing strategic capacity. *Financial Accountability and Management*, 15(3 and 4), 419–426.
- Pimm, S. L. (1984). The complexity and stability of ecosystems. *Nature*, 307, 321–326.
- Schattka, M., Puchkova, A., & McFarlane, D. (2016). Framework for simulation-based performance assessment and resilience improvement. *IFAC-PapersOnLine*, 49(12), 289–294.
- UNICEF. (2015) Annual report Guatemala. Guatemala: UNICEF. Retrieved from https://www.unicef.org/about/annualreport/files/Guatemala_2015_COAR.pdf.
- Vries, M., & Nemeč, J. (2013). Public sector reform: An overview of recent literature and research on NPM and alternative paths. *International Journal of Public Sector Management*, 26(1), 4–16.
- Walker, B., Gunderson, L. H., Knizig, A., Folke, C., & Carpenter, S. (2006). A handful of heuristics and some propositions for understanding resilience in social-ecological systems. *Ecology and Society*, 11(1), 13.
- Walker, B., Holling, C. S., Carpenter, S. R., & Kinzig, A. (2004). Resilience, adaptability and transformability in social-ecological systems. *Ecology and Society*, 9(2), 5.
- Walker, B., & Meyers, J. A. (2004). Thresholds in ecological and social-ecological systems: A developing database. *Ecology and Society*, 9(2), 3.
- Walker, B., & Salt, D. (2012). *Resilience thinking: Sustaining ecosystems and people in a changing world*. Washington, DC: Island Press.
- Wildavsky, A. B. (1988). *Searching for safety*. Brunswick: Transaction publishers.
- World Bank. (2003). *Poverty in Guatemala*. Washington, DC: The International Bank for Reconstruction and Development. Retrieved from <https://openknowledge.worldbank.org/bitstream/handle/10986/15066/275860PAPER0English0Poverty0in0Guatemala.pdf?sequence=1&isAllowed=y>.

Index

A

ANVUR guidelines, 282, 284

B

Benchmarking, 2, 3

Blame-avoidance, 46, 49, 50, 52, 54, 56, 57

C

Caesarean sections, 64

Case study, 2, 3

Causal mechanisms, 332, 333, 344

Contingent arguments, 4, 17, 18

Coordination, 162–166, 168, 170–175

Corruption, 252–256, 258–264

D

Data envelopment analysis, 2, 3

Dynamic performance management, 64, 69, 71, 73

Dynamic simulation, 22, 24, 25

E

Event management, 2, 3, 5, 7, 8, 14, 18

Exploratory case study, 52

G

Governance, 162, 164, 166, 173–175

Governance structure, 108, 110, 118

H

Health care, 1, 17, 25, 26, 28–40

Hierarchy, 123, 129, 163, 169, 173, 175

Higher education, 290–295, 302

I

Impacts, 204, 207, 212, 213, 215–220

Implementation, 179–183, 185–187, 198, 199

Indicators, 308, 311, 315, 316, 318, 319, 324, 326

Innovation, 203–208, 210–221

Inter-institutional cooperation, 14, 15

IPSAS RPG 3, 271, 277, 281, 283, 284

Italian public sector, 26, 38

Italy, 255, 256, 263, 270, 274, 275

J

Justice sector, 335, 343

L

Literature review, 205, 208–210, 212, 213, 217, 218

M

Maternity pathway, 64, 66, 74

Mathematical modeling, 5, 12, 24, 25

Multistakeholder engagement, 15

N

Network, 162–166, 168, 169, 171, 173–175

Network governance, 3, 14

New York State, 1, 2, 5, 7, 8, 17–19

Non-financial performance, 132

O

Organisational performance management, 82, 83, 92–95, 99–104

Organizational behaviour, 302

- Organizational performance, 27, 29, 31, 38, 39
 Outcome-based, 82–84, 100–104
 Outcome-based performance measurement systems, 4, 17
 Outcome indicators, 336, 339, 343, 344
 Outcome measures, 45, 46, 49–52, 54, 57, 64, 69, 72, 74
 Outcomes, 3, 6–16, 18
- P**
 Performance, 162, 164, 166, 172–175, 203, 204, 206–208, 210, 211, 215–221, 307–313, 315, 316, 318, 319, 324–326
 Performance-based funding, 290, 294–296, 301, 302
 Performance evaluation, 2, 15
 Performance management, 1, 5, 6, 11, 331, 333–338, 340–344
 Performance management use, 331, 334, 335, 338, 343, 344
 Performance measurement, 82, 83, 97, 100, 269–273, 276, 278–280, 283, 284
 Performance measurement system, 290, 295
 Performance measures, 107–111, 113–115, 117, 118
 Politicians, 3, 6, 8–10, 13, 16–18
 Public administrations, 252, 253, 255–258, 262, 263
 Public management, 245
 Public managers, 3, 6, 10, 12, 13, 16–18
 Public performance measurement., 308
 Public policy, 179, 180, 199
 Public policy analysis, 64, 69
 Public universities, 307, 308, 312–314, 316, 324–326
 Public value, 123, 130, 133–135
 Public value creation, 231
- R**
 Reporting, 308, 310, 311, 318, 319, 324
 Resilience, 1–9, 12–19
 Resource allocation, 290–292, 294–298, 300, 302, 303
- S**
 Scale of politico-administrative systems, 108, 109, 111, 112, 117, 118
 School districts, 1–8, 9, 12, 15–18
 Scotland performs, 82, 83, 85, 87, 93, 96, 99, 101, 102, 104
 Scottish government, 82–85, 87, 89–91, 93, 94, 96–99, 101–104
 Social mechanisms, 51
 Support services, 314–316, 318, 320, 324
 System dynamics, 5, 11, 12, 25, 179, 181, 199
- T**
 Training evaluation, 244
 Transparency, 252–258, 260–264
- U**
 Universities, 270–280, 282, 284, 285