Management for Professionals

Fiorenzo Franceschini
Maurizio Galetto
Domenico Maisano

Designing Performance Measurement Systems

Theory and Practice of Key Performance Indicators

Foreword by Andy Neely

Springer

Management for Professionals

More information about this series at http://www.springer.com/series/10101

Fiorenzo Franceschini • Maurizio Galetto •
Domenico Maisano

Designing Performance Measurement Systems

Theory and Practice of Key Performance Indicators

Foreword by Andy Neely

Springer

Fiorenzo Franceschini
Department of Management
and Production Engineering (DIGEP)
Politecnico di Torino
Turin, Italy

Maurizio Galetto
Department of Management
and Production Engineering (DIGEP)
Politecnico di Torino
Turin, Italy

Domenico Maisano
Department of Management
and Production Engineering (DIGEP)
Politecnico di Torino
Turin, Italy

ISSN 2192-8096 ISSN 2192-810X (electronic)
Management for Professionals
ISBN 978-3-030-01191-8 ISBN 978-3-030-01192-5 (eBook)
https://doi.org/10.1007/978-3-030-01192-5

Library of Congress Control Number: 2018957268

Cover illustration: eStudio Calamar, Berlin/Figueres

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

Foreword

Having been a student of performance measurement for over 25 years, it is a real treat to come across a book that sheds new light on a subject that has been so well explored. Numerous authors have tackled questions such as “How to design a performance measurement system?”, “How to ensure that metrics, behaviours and strategy are aligned?”, “What makes a good key performance indicator?”, “How and why should targets be set?” and “What role do incentives play in driving organizational performance?”.

Designing Performance Measurement Systems: Theory and Practice of Key Performance Indicators by Fiorenzo Franceschini, Maurizio Galetto and Domenico Maisano tackles many of these classic questions, but doesn’t rely on the standard treatments to do so. Instead, by combining historical insights with ideas from mathematics, quality and process engineering, this book does not simply shed new light on the subject, but rather it sheds new lights. The multiple perspectives and different ideas presented in the book will help you look at performance measurement in new and interesting ways.

The books consists of five chapters. The first explores the meaning of performance indicators, especially in the context of process performance. Exploring concepts such as the function and use of indicators, as well as different methods for classifying them, this introductory chapter sets out the breadth of material to be covered in the rest of the book. In the second chapter, Franceschini et al. look at uses and abuses of key performance indicators. Drawing on a wide range of examples, from life expectancy to air quality, they explore the strengths and weaknesses of different types of indicators. The third chapter draws on traditional measurement theory to develop “indicator theory”, arguing that a core concept is the condition of “non-uniqueness” for indicators. In the fourth chapter, Franceschini et al. turn their attention to properties of indicators, building on the ideas set out on indicator theory. Finally, the book closes with a chapter devoted to designing performance measurement systems, which covers some of the major reference models in use today, along with observations on the process of constructing performance measurement systems.

Over the years, I have read literally hundreds of books and papers on performance measurement. I have worked with thousands of managers, seeking to help them design and deploy better measurement systems. In *Designing Performance Measurement Systems: Theory and Practice of Key Performance Indicators*, Franceschini et al. introduce some new and interesting concepts which I am sure—like me—many students of performance measurement will find fascinating and rewarding.

Pro-Vice-Chancellor for Enterprise and Business Relations Andy Neely
University of Cambridge
Cambridge, UK
August 2018

Preface

In recent years, the use of indicators as a vehicle to transmit information, behavioural codes and rules of governance has dramatically increased. Indicators often accompany the daily life of public and private organizations in many fields: from the stock exchange to the meteorology and from the manufacturing processes to the sport specialties. Indicators regulate and influence organizations and their behaviour. In the broadest sense, they often give the impression to be the real driving force of social systems, economy and organizations.

The need to establish long-term objectives, rules and behaviours in order to achieve the planned results puts indicators in the spotlight of stakeholders of organizations.

Indicators take on the role of real "conceptual technologies", capable of driving organizational management in intangible terms, conditioning the "what" to focus and the "how"; in other words, they become the beating heart of the management, operational and technological processes.

Designing a performance measurement system is not an easy task. It requires a multidisciplinary approach that integrates knowledge from different process areas, people and information technologies and suitable scientific methods to ensure appropriate academic rigour.

This monograph describes in detail the main characteristics of indicators and performance measurement systems and summarizes methods and approaches for identifying, constructing and analysing indicators, combining theoretical and practical aspects.

The book is intended for academics, professionals and consultants involved in data analysis and indicator management. The description is relatively simple and does not necessarily require familiarity with advanced mathematics. The book can also be used in programmes for professionals, including senior executives, quality engineers, production engineers and procurement specialists. Professionals can also use the book for individual study.

This book is organized into five chapters. The first chapter deals with the basic concepts of indicators and process performance. The second chapter deals with the critical aspects, problems and curiosities that can arise when representing a generic system by means of indicators. The third chapter develops an original theory of indicators, showing that measurements can be seen as "special" indicators. The concept of “non-uniqueness of representation” by means of indicators will also be explained. Then the fourth chapter analyses in detail the properties of indicators. Description of the third and fourth chapter is supported by a number of practical examples and applications. Finally, the fifth chapter describes how performance measurement systems can be designed, implemented and maintained over time.

The content of this book is largely based on the monograph *Management by Measurement: Designing Key Indicators and Performance Measurement Systems*, produced by the same authors and published in 2007 by Springer-Verlag, Berlin Heidelberg.

Authors wish to gratefully acknowledge the support of all colleagues and friends who have contributed to the realization of the book with stimulating suggestions and helpful comments, including Emil Bashkansky, Luca Mastrogiacomo and Sergio Rossetto.

Turin, Italy
July 2018

Fiorenzo Franceschini
Maurizio Galetto
Domenico Maisano

Contents

1 Quality Management and Process Indicators

Abstract

This chapter introduces the problem of constructing a *quality management system*, i.e., a system aimed at suitably driving and controlling an organization. To this purpose, it is essential to (1) identify the most characteristic processes of the organization of interest; and (2) monitor and evaluate them regularly. This demonstrates the important role played by indicators even at a "normative" level.

This chapter clarifies these general concepts, referring to recent international standards about quality-management principles and methods. It then shows a preliminary classification of indicators and a description of their general characteristics. Finally, an overview of the main state-of-art research fronts on indicators is provided.

1.1 General Concepts

Complex organizations implement performance measurement systems in order to give due attention to results, responsibilities and targets. For example, knowing the performance in terms of sales and customer satisfaction allows a manufacturing company to "feel the pulse" of the market and plan its future development. Managers utilize indicators to allocate assets or to make decisions on the best strategies.

While *quality* standards have become central operational tools for organizations, performance indicators are the *communication protocol* of their health state to the outside world. An extensive empirical research, carried out in the United States, showed that the organizations winning quality awards are usually those with higher profits (Hendricks and Singhal 1997).

But how can we recognize the quality of organizations? Paraphrasing the standard ISO 9000:2015 (2015), quality is the ability to fulfil different types of requirements—e.g., productive, economical, social ones—with tangible and measurable actions. Quality is a basic element to differentiate an organization with respect to its

F. Franceschini et al., *Designing Performance Measurement Systems*, Management for Professionals, https://doi.org/10.1007/978-3-030-01192-5_1

competitors. To make quality tangible, it is firstly necessary to identify stakeholders' needs. Then it is necessary to fulfil these needs effectively, using the available assets (i.e., processes and resources). This requires a careful analysis if the evolution of processes; performance indicators are suitable tools to achieve this purpose.

Indicators are not just "passive" observation tools but can have a deep "normative" effect, which can modify behaviour of organizations and influence decisions. If a production-line manager is trained to classify the products that are spread over the market as "good", his/her attention will be directed towards maximizing the diffusion and expansion of these products; unintentionally, this strategy could sacrifice long-term profits or investments in other products. If a *call center* administrator is recompensed depending on his/her ability to reduce absenteeism, he/she will try to reach this target even if that will not necessarily lead to increase productivity.

The mechanism is easy to work out, as exemplified by Hauser and Katz (1998): if an organization measures indicators a, b and c, neglecting x, y and z, then managers will pay more attention to the first ones. Soon those managers who do well on indicators a, b and c are promoted or are given more responsibilities. Increased pay and bonuses follow. Recognizing these rewards, managers start asking their employees to make decisions and take actions that improve these indicators and so on. The organization gains core strengths in producing a, b and c. Organizations become what they measure! (Hauser and Katz 1998).

If maximizing a, b and c leads to long-term profit, the indicators are effective. If a, b and c lead to counterproductive decisions and actions, then indicators have failed. Even worse, once the organization is committed to these indicators, indicators provide tremendous inertia. Those who know how to maximize a, b and c fear to change the course, since it is generally very difficult to refocus an organization on new goals.

Unfortunately, selecting good indicators is not so easy. This book focuses on the construction of performance measurement systems, being aware that "magic rules" to identify them do not exist. Many indicators seem right and are easy to measure but may have counter-productive consequences. Other indicators are more difficult to measure but they may address the organization to those decisions and actions that are critical to success.

This book will suggest how to identify indicators that enhance long-term profitability, consistently with the goals of quality management. First, it is necessary to identify stakeholders' exigencies; then, it is necessary (1) to define performance levels, (2) organize and control the activities involved in meeting the targets (practices, tasks, functions), (3) select indicators, (4) define how to gather information, and (5) determine corrective or improving actions.

1.2 Quality Management Systems

A *quality management system* is a set of tools for driving and controlling an organization, considering all different quality aspects (ISO-9000:2015 2015):

- human resources;
- know-how and technology;
- working practices, methodologies and procedures.

A quality management system should accomplish specific planned targets such as production, cost, time, return of investment, stakeholders exigencies or expectations, supporting the following operations:

- performance evaluation of several organizational aspects (processes, suppliers, employees, customer satisfaction, etc.);
- market analysis (shares, development opportunities, etc.);
- productivity and competitor analysis;
- decisions about product innovation or new services provided.

For achieving positive results on many fronts (market shares, productivity, profit, competitiveness, customer portfolio, etc.), it is essential that organizations implement quality-management principles and methods.

According to the ISO 9000:2015 (2015) standard, the creation of quality management systems is supported by seven fundamental principles:

1. *Customer focus*. Organizations must understand the customer needs, requirements and expectations.
2. *Leadership*. Leaders must establish a unity of purpose and set the direction that an organization should follow. Furthermore, they must create the conditions for people to achieve the objectives.
3. *Engagement of people*. Organizations must encourage the commitment of employees and the development of their potential at all hierarchical levels.
4. *Process approach*. Organizations are more efficient and effective when adopting a process approach to manage activities and related resources. This approach must also be systemic, i.e. interrelated processes should be identified and treated as a system.
5. *Improvement*. Organizations must be encouraged to continuously improve their performance.
6. *Evidence-based decision making*. Strategic decisions should rely on the analysis of factual data.
7. *Relationship management*. Organizations must maintain a mutually beneficial relationship with interested parties (e.g., suppliers, service providers, third parties, etc.) so as to help them create value.

These principles should be applied to improve organizational performance and achieve success. The main advantages are:

- Benefits concerned with marketing and customer relationships:
  - support for development of new products;
  - easier access to market;

  - customers are aware of research and quality efforts by organizations;
  - better credibility of organizations.
- Internal benefits:
  - quality is easier to plan and control;
  - support for the definition of internal standards and work practices;
  - more effective and efficient operations.
- Benefits concerning relationships with interested parties:
  - better integration with interested parties;
  - reduction of the number of suppliers and use of rational methods for their selection and evaluation;
  - increased capability to create value for interested parties, by sharing resources/competences and managing quality-related risks.

1.3 The Concept of Process

1.3.1 Definition

According to the ISO 9000:2015 (2015) standard, a process is "*a set of interrelated or interacting activities that use inputs to deliver an intended result (output)*". This general definition identifies the process like a black box, in which input elements are transformed into output ones.

The process approach is a powerful management tool. A system is generally made of several interconnected processes: the output of one process becomes the input of one other, and so on. Processes are "glued" together by means of such input-output relationships. When analysing each process, it is necessary to identify the target characteristics of the output, and the so-called stakeholders; not only final users, but all the parties involved in the process—inside and outside the organization—should be considered.

Increasing the level of detail of the analysis, each process can be decomposed into sub-processes, and so on. This sort of "explosion" should be reiterated, in order to identify all the basic components of the organization.

Monitoring a process requires identifying specific activities, responsibilities and indicators for testing effectiveness and efficiency. *Effectiveness* means setting the right goals and objectives, making sure that they are properly accomplished (*doing the right things*); effectiveness is measured comparing the achieved results with target objectives. On the other hand, *efficiency* means getting the most (output) from the available (input) resources (*doing things right*): efficiency defines a link between process performance and available resources.

1.3.2 Process Modeling

To manage processes, we need a proper modeling, which considers major activities, decision-making practices, interactions, constraints and resources. It is important to identify the relevant process characteristics and then represent them.

Modeling a process means describing it, never forgetting the targets which should be met. Process is a symbolic "place" where customer expectations are turned into organizational targets, and these targets are turned into operative responses. A proper performance measurement system should be set up to verify the consistency of responses with requirements.

The goal of process modeling is to highlight process characteristics and peculiarities (e.g., organizational, technological, relational aspects, etc.). Process modelling is generally supported by software applications, which map and display activities/actors involved, focusing on several process aspects (input, output, responsibilities, etc.) and practical parameters (time, cost, constraints, etc.).

Mapping is essential to understand the process. It is possible to perform process performance simulations, identifying "optimal" operational conditions, in terms of costs, time and quality. A significant support to managers is given by process representation tools, such as IDEF, CIMOSA, DSM, etc. (CIMOSA 1993; Draft Federal Information 1993; Mayer et al. 1995; Ulrich and Eppinger 2000; Li and Chen 2009). These methodologies make it possible to manage different perspectives of the organization: functions, activities, resources and physical/informative flows.

1.3.3 Process Evaluation

Since the object of a generic process is meeting stakeholder needs, this condition has to be evaluated through suitable process measures. To this purpose, evaluating the performance/evolution of processes is essential.

According to the UNI 11097:2003 standard (UNI-11097 2003): "*A system of indicators should become an information system for estimating the level of achievement of quality targets*".

Indicators should be selected considering:

- quality policy;
- quality targets;
- the area of interest within the organization, e.g., market competitiveness, customer satisfaction, market share, economical/financial results, quality, reliability, service level, flexibility of service supply, research and development, progress and innovation, management, development and enhancement of human resources, internal and external communication;
- performance factors;
- process targets.

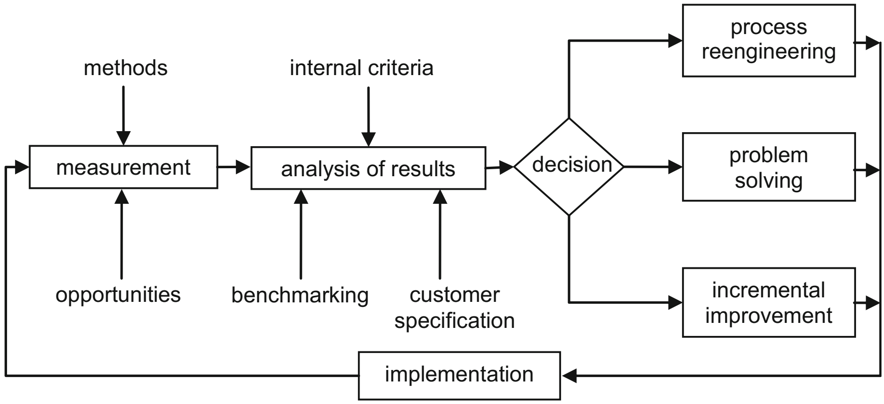

Fig. 1.1 The process improvement chain (Barbarino 2001). With permission

It should be remarked that any deficiency in the system for evaluating the performance of a process will affect the so-called *non quality* costs. These costs represent a powerful and rational lever to persuade organizations to improve continuously.

Process implementation should be followed by a systematic monitoring plan and periodical performance recording, in order to identify critical aspects and/or reengineer process activities. Figure 1.1 represents this concept.

A system for evaluating the performance of a process generally requires two activities:

1. *Definition of indicators*. This phase concerns the definition of the indicators to use and relevant data to collect. Indicators are selected depending on the critical aspects and growth potential of the process.
2. *Decision*. Depending on the difference between target and measured performance level, there are three different courses of action:
   – individual problem solving;
   – incremental improvement (*step by step*);
   – process reengineering.

As represented by the feedback loop in Fig. 1.1, a performance measurement system includes a "self-regulating" mechanism. Output data of the process are used as input data for the performance measurement system, in order to drive possible actions or decisions. A crucial point of this approach is the implementation of the performance measurement system.

The organization management is the final receiver of process-monitoring activities, whose results are used to make decisions concerning the allocation of resources and responsibilities. These decisions may influence the future behaviour of the organization.

Evaluations should be technically and economically efficient and should focus on results instead of actions. For example, quantitative variables are generally more practical than qualitative ones: while the former ones can be easily referred to monetary values, the latter ones are more practical in describing the organization behaviour and the consequences of past actions.

1.4 Process Indicators

As explained before, measuring is essential for the process-performance control and improvement. However, constructing and implementing a measurement system is easier said than done. The crucial point is to identify the "right" indicators to properly represent the process: i.e., the so-called *Key Performance Indicators* (KPIs) (Petersen et al. 2009).

The UNI 11097 (2003) standard classifies as quality indicator *"the qualitative and/or quantitative information on an examined phenomenon (or a process or a result), which makes it possible to analyze its evolution and to check whether quality targets are met, driving actions and decisions"*.

Several crucial points in the construction of indicators are: (1) they should appropriately represent the process of interest; (2) they should be well-understood and accepted by process managers and employees; (3) they should be traceable and verifiable.

Generally, each indicator refers to a specific *target*, which can be seen as a reference for comparisons. This reference can be *absolute* or *relative* (e.g., depending on whether it is external or internal to the organization). A "zero-defects" program is an example of absolute reference. Reference values can be derived from the organization's past experience or even extrapolated from similar processes (benchmarking).

The indicator definition by UNI 11097 (2003) entails some basic requirements:

- indicators should represent targets effectively;
- they should be simple and easy to interpret;
- they should be able to indicate time trends;
- they should "respond" to changes within or outside the organization;
- the relevant data collection and data processing should be easy;
- they should be updated easily and quickly.

Relevant characteristics and properties of indicators will be discussed on Chap. 4.

Quality factors (or *dimensions*) are the most significant aspects for characterizing the state of a process. Each of them should be identified and associated to one or more process indicator(s).

The UNI 11097 (2003) standard explains that "*measurements of the examined phenomenon should be faithfully and properly documented, without any distortion or manipulation. The information provided by indicators should be exact, precise and responsive to significant changes, as well as reproducible.*"

One of the most difficult activities in process management is making the process performance "tangible". Process managers try to do this, translating organization goals into different metrics/indicators, which are also visible from the outside world. This *modus operandi* can be applied to any type of process: a manufacturing process, a service, or a generic organization. The typical question addressed by process managers is: "Does process performance meet the expected targets?" (Magretta and Stone 2002).

When translating an organization's mission/strategy into reality, choosing the right indicators is a critical aspect. Indicators and strategies are tightly and inevitably linked to each other: a strategy without indicators is useless, while indicators without a strategy are meaningless (Franco-Santos et al. 2012).

The interest towards indicators is increasing and their importance has been long recognized in several contexts. Every organization, activity or worker needs indicators as they drive the activities of measuring (i.e., evaluating how we are doing), educating (i.e., what we measure indicates the way we plan to deliver value to our customers), and directing (i.e., potential problems are related to the gaps between indicators and targets). This book tries to highlight the potential of indicators, as well as their drawbacks.

Yet, indicators continue to be a challenge to managers and researchers. While there are numerous examples of indicators (e.g., in the field of Logistics, Quality, Information Sciences, System Engineering, Sustainable Development), there are relatively few studies focused on their development (Rametsteiner et al. 2011). Some examples can be found in the research of Beaumon (1999), Leong and Ward (1995), Neely (1998, 2002, 2005, 2007), New and Szwejczewski (1995), and Bourne et al. (2003). A great deal of what we currently know about indicators comes from the managerial literature, e.g. (Brown 1996; Dixon et al. 1990; Kaydos 1999; Ling and Goddard 1988; Lockamy and Spencer 1998; Lynch and Cross 1995; Maskell 1991; Melnyk and Christensen 2000; Neely and Bourne 2000; Neely et al. 2000; Smith 2000; Choong 2014).

The perspective of managers differs from that of researchers, due to their different priorities. Researchers are generally concerned with defining, adapting and validating indicators to address specific research questions. The time required to develop and collect indicators is less important than the validity and generalizability of the results beyond the original context. On the other hand, managers face far greater time pressure and are less concerned about generalizability. They are generally willing to use a "good enough" indicator, if it can provide useful information quickly. However, as long as the difference in priorities is recognized, the two points of view are gradually becoming closer. Undoubtedly, researchers can contribute to managers' understanding of indicators, while the managers' can help researchers in studying the practical impact of indicators and measuring procedures (Perkmann et al. 2011).

Recent studies suggest that indicators are receiving more and more attention, due to their strategic role for process management; many research programs all over the world have been dealing with these questions. For example, KMPG (i.e., an international private company) in collaboration with the University of Illinois undertook a major research focused on performance measurement (funding of about US $3 millions).

The January 2003 *Harvard Business Review* case study focused on the miscues and disincentives created by poorly thought out performance measurement systems (Kerr 2003). What are the reasons of this increasing interest on indicators? Here are some possible reasons:

- "never satisfied" consumers (McKenna 1997);
- the need to manage the "total" supply chain, rather than internal factors separately (holistic vision);
- shrinking of products/services life cycle;
- bigger and bigger (but not necessarily better) data;
- an increasing number of decision-support tools which utilize indicators.

The above reasons stimulate the construction of new performance indicators and approaches, which allow to identify improvement opportunities and anticipate potential problems (Smith and Bititci 2017). Additionally, indicators should be considered as important tools to identify and share the priorities of organizations across the supply chain. In fact, indicators misalignment is thought to be a primary source of inefficiency and disruption in supply-chain interaction.

1.4.1 Functions of Indicators

Indicators represent a way of "distilling" the larger volume of data collected by organizations. As data become bigger and bigger, due to the greater span of control or growing complexity of operations, data management becomes increasingly difficult. Actions and decisions are greatly influenced by the nature, use and time horizon (e.g., short or long-term) of indicators.

Indicators provide the following three basic functions:

- *Control*. Indicators enable managers and workers to evaluate and control the performance of the resources that they are supposed to manage.
- *Communication*. Indicators communicate performance to internal workers and managers, and to external stakeholders too. On the contrary, incomplete/inappropriate indicators may produce frustration and confusion.
- *Improvement*. Indicators identify gaps (between performance and targets) that ideally point the way for possible improving actions. The size of these gaps and their direction (e.g., positive or negative) can be used to adjust/plan corrective actions.

Each system of indicators is subject to a dynamic tension, which stems from the desire to introduce new changes in response to new strategic priorities, and the desire to maintain "old" indicators to allow comparison of performance over time. This tension will determine the so-called *life cycle* of indicators.

Fig. 1.2 Classification of indicators on the basis of focus and tense attributes (Melnyk et al. 2004). With permission

INDICATOR FOCUS \ INDICATOR TENSE	Outcome	Predictive
Financial	Return on assets	Time financial flow
Operational	Flow times	No. of process steps and setups

1.4.2 Aims and Use of Indicators

Regarding indicators, one source of complexity is their great variety. Various indicators can be classified according to two attributes: indicator focus and indicator tense.

Indicator focus pertains to the resource that is the focus of the indicator *per se*. Generally, indicators report data in either financial (monetary) or operational terms (e.g., operational details such as lead times, inventory levels or setup times). Financial indicators define the pertinent elements in terms of monetary resource equivalents, whereas operational indicators tend to define elements in terms of other resources (e.g., time, people) or outputs (e.g., physical units, defects).

The second attribute, *indicator tense*, refers to how indicators are intended to be used. Indicators can be used both to judge outcome performance (*ex post*) and to predict future performance (*ex ante*). Many of the cost-based indicators used in organizations belong to the first category. In contrast, a predictive use of an indicator is aimed at increasing the chances of achieving a certain objective. If our interest is to reduce *lead time*, then we might use indicators like the “setup time” or the “number of process steps”. The emphasis on the identification and use of indicators in a predictive way is relatively new. Predictive indicators are appropriate when the main interest is preventing the occurrence of problems, rather than correcting them.

The combination of these attributes (focus and tense) provides four distinct categories of indicators, as shown in Fig. 1.2. Top managers are generally more interested in financial/outcome indicators. In contrast, operations managers and workers are generally more interested in operational, predictive or outcome indicators.

1.4.3 Terminology

The terminology used in the performance measurement context is not completely and univocally defined. Similar concepts are often classified using different terms, depending on the technical area of interest. For example, terms as “performance metric”, “measure” and “indicator” are usually considered as synonyms. The same happens for terms such as “target”, “result” or “performance reference”. In the following descriptions we try to clarify the meaning of each of these terms. The proposed terminology will generally conform to the ISO 9000:2015 (2015) standard.

1.4.4 Categories of Indicators

The term "indicator" is often used to refer to one of the following categories (which are partially linked to each other, as explained in Chaps. 3 and 4):

- *basic* indicators;
- *derived* indicators;
- *sets* of indicators;
- whole performance measurement systems.

Basic indicators are like "elementary elements" that can be aggregated into *derived* indicators, which—in turn—represent the synthesis of (sub)indicators. A *set of indicators* is supposed to represent a specific process function.

A *performance measurement system* coordinates the indicators related to the various functions, from the strategic level (top management) to the operational level (shop floor/purchasing/execution context). For a certain activity/product/function, multiple indicators can be developed and implemented. The challenge is to design an organic framework of indicators, which makes it possible to depict the overall performance of a process from several perspectives.

In the current literature, several different approaches have been proposed, including:

- the *Balanced Scorecard* method (Kaplan and Norton 1992, 1996, 2001, 2008; Ittner and Larcker 1998; Bhagwat and Sharma 2007; Taylor and Baines 2012);
- the *Strategic Profit Impact* model, also known as the *Dupont* model (Lambert and Burduroglu 2000; Stapleton et al. 2002);
- the *Critical Few* method (U.S. Department of Energy - PBM SIG 2012);
- the *EFQM (European Foundation for Quality Management)*, and *Malcom Baldrige Quality Award* model (EFQM 2013; BNQP 2018; NIST 2018).

Each of these approaches has strengths and weaknesses, which will be discussed in detail in Chap. 5. For example, the *Balanced Scorecard* forces top management to recognize that multiple activities should be carried out for the success of the organization. Management of these activities must be balanced: all the organization's features (dimensions) should be considered, not only the economical ones. Furthermore, this approach gives useful information on how to perform a practical synthesis of the most relevant indicators.

The performance measurement system is ultimately responsible for maintaining *alignment* and *coordination*. Alignment deals with the maintenance of consistency between strategic goals and indicators. Alignment means that objectives, which are set at higher levels, should be consistent with indicators and activities at lower levels.

In contrast, coordination recognizes the presence of interdependency between processes, activities or functions. Coordination deals with the degree to which indicators related to various areas are consistent and/or complementary with each other. Coordination strives to reduce potential conflict when there are contrasting goals. E.g., in the manufacturing field, productivity indicators (number of elements produced) conflict with quality indicators (number of defects).

A good set of indicators directs and regulates the activities consistently with strategic objectives and provides real-time feedback, predictive data, and insights into opportunities for improvement. In addition, indicators need to be responsive to changes (e.g., in process conditions, input, resources, etc.).

The coming chapters will deepen these themes, with special attention to the potential and limitations of indicators.

1.4.5 A General Classification of Indicators

Indicators should provide accurate information about the evolution of a process. UNI 11097 (2003) standard suggests an interesting classification, which depends on the "observation moment" of a process.

There are three types of indicators, which are individually discussed in the three following subsections:

- *Initial* (or structure) indicators. For example, indicators concerned with the quality of materials or quality of services provided by suppliers.
- *Intermediate* (or process) indicators. For example, indicators related to the efficiency of a manufacturing process.
- *Final* (or result) indicators. For example, indicators of customer satisfaction or production cost.

Initial Indicators (or Structure Indicators)

Planning is the first task in a project as it makes it possible to evaluate the ability of one organization to meet its targets, considering the available (organizational, physical and monetary) resources.

Initial indicators—or structure indicators—are supposed to answer the question: "What are the available assets and the working patterns of the process, considering all the resources involved (e.g., facilities, human resources, technological and monetary assets, services provided by suppliers, and so on)?"

These indicators are also used to qualify the skill and involvement level of human resources, with the final purpose of improving the project planning/management.

Intermediate Indicators (or Process Indicators)

Intermediate indicators are supposed to answer the question "How does the process work?" They measure the consistency between process results and process specifications, providing useful information on the process state. This type of control makes it possible to understand whether process conditions are stable or whether process has run into unexpected or unpredictable difficulties.

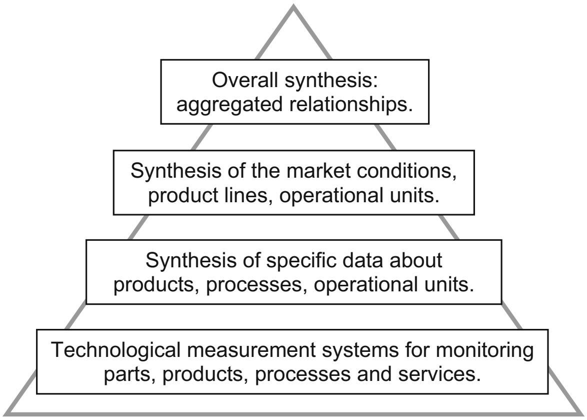

Fig. 1.3 The measurement systems pyramid (Juran 2016). With permission

Final Indicators (or Result Indicators)

Final indicators (or result indicators) answer the following questions:

- “What are process outcomes?”;
- “Has the process met the purposes?”;
- “What are the expected/unexpected effects produced by the process?”;
- “What is the cost-benefit ratio?”.

Final indicators are generally viewed as the most important ones, since they are supposed to represent the final results of the process, both the positive and the negative ones. For example, they may deal with customer satisfaction or cost of products/services.

Another classification is based on the “position” of indicators within the organizational framework. Figure 1.3 represents the pyramidal categorization suggested by Juran (2016). At the pyramid bottom, there are “technological measurement systems” for monitoring (parts of) products, processes and services. At second level, indicators synthesise basic data on individual product or process: for instance the percentage of defects in a specific product or service.

Third level includes “quality measurement systems”, dealing with entire sectors, e.g., production lines or services. At the top of the pyramid, we find the “overall synthesis indicators”, which are used by top management to evaluate the conditions of economic/monetary aspects, manufacturing processes and market.

1.4.6 Comparison of Economic and Process Indicators

Economic-Financial Indicators

Process performances of each organization can hardly be monitored without using “monetary” indicators. Indicators derived from the general accounting are traditionally used to measure performances. These indicators are often easy to measure and user-friendly.

Since economic outcomes are the result of past decisions, economic indicators cannot be used to identify future opportunities of organizations. Classically, the most common drawbacks of economical/financial performance indicators are:

- They are not prompt; in fact, the evaluation and aggregation of physical transactions may require a lot of time (especially for organizations with a large range of products);
- They generally report information to the outside, rather than to the inside of the organization;
- They focus on costs, in particular labour costs, which are nowadays less decisive in the determination of the surplus value of processes;
- They ignore quality, innovation potential, competencies, skills improvement, and the strategic dimensions of competitiveness and added-value;
- Sometimes they slow down the development of new and more suitable (than the existing) organizational structures;
- They are not very sensitive to changes in the organization's strategies, or in the external context.

Reduced timeliness is the major limitation of financial indicators. To determine these indicators, the information of interest (e.g., market shares, product and process characteristics, etc.) needs to be translated into monetary terms. While economic indicators can be calculated from the final balance, financial indicators need to estimate future results. Consequently, they require a more complex and frequent data collection, so as to identify problems promptly.

One of financial indicators strengths is long-term orientation, which derives from the joint analysis of short-term and long-term results.

Considering financial indicators, the link between their exhaustiveness and precision depends on the analysis depth. The more strategic aspects are examined in detail, the more the analysis will result complete.

The use of general accounting indicators should be limited to organizations operating within a stable context, where short-time profitability may properly represent competitiveness. When the context is dynamic, it is more difficult to identify a correlation between past and future results and therefore the information needs to be more timeliness (Bititci et al. 2006).

There are two possible solutions to overcome these limitations: (1) improving the current financial indicators and/or (2) focusing on the operational measurements. As a matter of fact, managers' analysis is based on more than one indicator, in order to cover all the critical aspects for business. To this purpose, it is required to construct a balanced representation of both financial and operational measurements (Kaplan and Norton 1996, 2001, 2008).

Process Indicators

Process indicators can be classified depending on the measured *competitiveness factor* (time, quality, flexibility, productivity and environmental compatibility) and their purpose (indicators of customer satisfaction or indicators to manage internal resources).

Time indicators typically relate to the time for product development or logistic development process. Time indicators can be divided into two main categories. The first sees time as a source of internal efficiency; in this case, saving time means cost reduction and creation of value. The second category includes those indicators related to timeliness in market response. In this case, time is viewed as a lever for the product differentiation. Internal (time) indicators aim at identifying those process activities which produce added-value, such as those for improving the customer perception of a certain product/service. On the other hand, external (time) indicators can be divided into the following families:

- *Delivery timeliness of standard products*. E.g., indicators aimed at evaluating the competitiveness of logistic systems;
- *Development time of new products*. E.g., indicators aimed at evaluating the competitiveness of the products development process. The most common indicator is "time to market", i.e., time period between concept definition and market launch of the product.

Quality indicators investigate the product/service characteristics compared to customer needs (compliance with product specifications and customer needs) and process efficiency/effectiveness (resource waste, defectiveness, etc.).

Productivity indicators are represented by classical process indicators, which are generally defined as the ratio of process outputs to process inputs and are used to estimate labour productivity. Productivity indicators are typically used in manufacturing industries, where output results can be easily measured.

Indicators of environmental compatibility aim at identifying the organization's ability to develop environmentally friendly products/processes. Although, in the past, environmental issues were confined to the technical/operational field, they are become more and more important for organizations.

Flexibility indicators evaluate the organization's ability to quickly respond to changes, keeping time and cost low. Consequently, the more the context is dynamic, the more flexibility is important. There are two typologies of changes: quantitative changes—related to positive or negative fluctuations in product/service demand—and qualitative changes—related to modifications in product/service typologies. Depending on the type and size of changes, we can identify six possible dimensions: volume, mix, modularity, product, production and operation.

The major distinctive element of process indicators is *timeliness*. While economic indicators entail that physical operations are translated into monetary terms, process indicators simply derive the information from transactions.

A second distinctive element is *long-term orientation*. Process indicators may provide a synthesis of the organization's competitive advantages.

Furthermore, it can be difficult to assess the *exhaustiveness* of process indicators. While economic/financial indicators generally aggregate several types of performance into a single monetary variable, process indicators are generally related to specific types of performance: e.g., a competitive time-to-market does not guarantee that the product quality will satisfy customers. A limitation of some indicators is the loss of sight of the organization's complexity (Bititci et al. 2006).

1.4.7 Indicators and Research: State of the Art

A "hot" front of the research on performance indicators is the study of their impact on complex systems (Bourne et al. 2005; Bititci et al. 2012). However, this topic is not completely new. Skinner (1974) identified simplistic performance evaluation as being one of the major causes for organizations getting into trouble. Subsequently, Hill (1999) recognized the role and impact of performance measures and performance measurements systems in his studies of manufacturing strategy. In these (and other) studies, indicators are often viewed as part of the infrastructure or environment in which manufacturing must operate (*conceptual technologies*).

However, there is still a need to allocate the topic of indicators into a theoretical context—i.e., a framework that gives a central role to indicators. An interesting theoretical framework for research is *agency theory* (Eisenhardt 1989). This theory applies to the study of problems arising when one party, i.e., the principal, delegates work to another party, i.e., the agent. The unit of analysis is the metaphor of a contract between the agent and the principal. What makes agency theory so attractive is the recognition that in most organizations the concept of contract as a motivating and control mechanism is not really appropriate. Following this idea, the contract is replaced by the indicator: it is the indicator that motivates and directs; it is the indicator that enables principals to manage and direct the activities of agents (Austin 1996).

Another interesting framework for future research is *dependency theory* (Pfeffer and Salancik 1978). This theory states that the degree of interdependence and the nature of interactions among functional specialists in an organization are influenced by the nature of the collective task that they seek to accomplish. In dynamic environments with rapid product changes and heterogeneous customer requests, agents should manage various tasks within several interconnected functional areas. Dependency theory has implications for the design of indicators systems. For example, it can be helpful to answer questions like: "How should indicators reflect the interdependencies of different functional areas?", or "How often should indicators be changed or adjusted?"

A third way to look at indicators is offered by Galbraith (1973). The basic idea is that, presumably, a richer set of indicators encourages communication among decision makers, workers, strategy representatives, and customer of a generic process. However, there may be limits to the organization's (as well as individuals') ability to process larger sets of indicators. Increasing numbers of indicators could lead to greater conflict and uncertainty regarding future actions. Given this apparent trade-off between richness and complexity of a set of indicators, an information-processing theoretical view could stimulate research into questions regarding the optimal size of an indicator set, or perhaps the optimal combination of outcome and predictive indicators included in the set.

An additional research issue is concerned with the verification of the *condition of uniqueness*: "Given a process, is there a unique set of indicators that properly represent it?" Chap. 3 will provide an answer to this other question.

Finally, other possible research topics include (Melnyk et al. 2015; Neely et al. 2005; Taticchi et al. 2012):

- evaluating the relationship between financial and operating indicators;
- measuring performance within a supply-chain environment;
- assessing consistency of indicators with respect to strategic plans;
- implementing *dynamic* (with respect to time) performance measurement systems;
- integrating performance measurements and rewards/incentives.

References

Austin, R. B. (1996). *Measuring and managing performance in organizations*. New York: Dorset House Publishing.

Barbarino, F. (2001). UNI EN ISO 9001:2000. Qualità, sistema di gestione per la qualità e certificazione, Il Sole 24 Ore, Milan.

Beaumon, B. M. (1999). Measuring supply chain performance. *International Journal of Operations and Production Management, 19*(3), 275–292.

Bhagwat, R., & Sharma, M. K. (2007). Performance measurement of supply chain management: A balanced scorecard approach. *Computers & Industrial Engineering, 53*(1), 43–62.

Bititci, U. S., Garengo, P., Dörfler, V., & Nudurupati, S. (2012). Performance measurement: Challenges for tomorrow. *International Journal of Management Reviews, 14*(3), 305–327.

Bititci, U. S., Mendibil, K., Nudurupati, S., Garengo, P., & Turner, T. (2006). Dynamics of performance measurement and organisational culture. *International Journal of Operations and Production Management, 26*(12), 1325–1350.

BNQP (Baldrige National Quality Program). (2018). *Criteria for performance excellence*. Gaithersburg, MD: NIST Publications. Retrieved September 2018, from http://www.baldrige.nist.gov.

Bourne, M. C., Kennerley, M., & Franco-Santos, M. (2005). Managing through measures: A study impact on performance. *Journal of Manufacturing and Technology Management, 16*(4), 373–395.

Bourne, M., Neely, A., Mills, J., & Platts, K. (2003). Implementing performance measurement systems: A literature review. *International Journal of Business Performance Management, 5*(1), 1–24.

Brown, M. G. (1996). *Keeping score: Using the right metrics to drive world-class performance*. New York: Quality Resources, CRC Press.

Choong, K. K. (2014). Has this large number of performance measurement publications contributed to its understanding? A systematic review for research and applications. *International Journal of Production Research, 52*(14), 4174–4197.

CIMOSA. (1993). *Open system architecture for CIM*. Edition Esprit, Consortium AMICE Editor. Berlin: Springer. Retrieved September 2018, from www.cimosa.de.

Dixon, J. R., Nanni Jr., A. J., & Vollmann, T. E. (1990). *The new performance challenge: Measuring operations for world-class competition*. Homewood, IL: Dow Jones-Irwin.

Draft Federal Information. (1993). *Announcing the Standard for "Integration Definition For Function Modelling (IDEF0)"*. Processing Standards Publication 183. Retrieved September 2018, from www.idef.com.

EFQM (European Foundation for Quality Management). (2013). Retrieved September 2018, from www.efqm.org.

Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *The Academy of Management Review, 14*(1), 57–74.

Franco-Santos, M., Lucianetti, L., & Bourne, M. (2012). Contemporary performance measurement systems: A review of their consequences and a framework for research. *Management Accounting Research, 23*(1), 79–119.

Galbraith, J. R. (1973). *Designing complex organizations*. Reading, MA: Addison-Wesley.

Hauser, J., & Katz, G. (1998). Metrics: You are what you measure! *European Management Journal, 16*(5), 517–528.

Hendricks, K. B., & Singhal, V. R. (1997). Does implementing an effective TQM program actually improve operating performance? Empirical evidence from firms that have won quality awards. *Management Science, 43*(9), 1258–1274.

Hill, T. (1999). *Manufacturing strategy: Text and cases*. Burr Ridge, IL: McGraw-Hill.

ISO 9000:2015. (2015). *Quality management systems – Fundamentals and vocabulary*. Geneva: ISO.

Ittner, C. D., & Larcker, D. F. (1998). Innovations in performance measurement: Trends and research implications. *Journal of Management Accounting Research, 10*, 205–238.

Juran, J. M. (2016). *Quality control handbook* (7th ed.). New York: McGraw-Hill.

Kaplan, R. S., & Norton, D. (1992). The balanced scorecard-measures that drive performance. *Harvard Business Review, January–February*, 71–79.

Kaplan, R. S., & Norton, D. (1996). *The balanced scorecard*. Cambridge, MA: Harvard Business School Press.

Kaplan, R. S., & Norton, D. P. (2001). *The strategy-focused organization: How balanced scorecard companies thrive in the new business environment*. Boston: Harvard Business School Press.

Kaplan, R. S., & Norton, D. P. (2008). *The execution premium. Linking strategy to operations for competitive advantage*. Boston: Harvard Business School.

Kaydos, W. (1999). *Operational performance measurement: Increasing total productivity*. Boca Raton, FL: St. Lucie Press.

Kerr, S. (2003). The best-laid incentive plans. *Harvard Business Review, January*, 27–37.

Lambert, D. M., & Burduroglu, R. (2000). Measuring and selling the value of logistics. *International Journal of Logistics Research, 11*(1), 1–17.

Leong, G. K., & Ward, P. T. (1995). The six Ps of manufacturing strategy. *International Journal of Operations and Production Management, 15*(12), 32–45.

Li, Q., & Chen, Y. L. (2009). IDEF3 process capture method. In *Modeling and analysis of enterprise and information systems*. Berlin: Springer.

Ling, R. C., & Goddard, W. E. (1988). *Orchestrating success: Improve control of the business with sales and operations planning*. New York: Wiley.

Lockamy III, A., & Spencer, M. S. (1998). Performance measurement in a theory of constraints environment. *International Journal of Production Research, 36*(8), 2045–20060.

Lynch, R. L., & Cross, K. F. (1995). *Measure up! How to measure corporate performance*. Malden, MA: Blackwell.

Magretta, J., & Stone, N. (2002). *What management is: How it works and why it's everyone's business*. New York: Free Press.

Maskell, B. H. (1991). *Performance measurement for world class manufacturing*. Cambridge, MA: Productivity Press.

Mayer, R. J., Menzel, C. P., Painter, M. K., De Witte, P. S., Blinn, T., & Perakath, B. (1995). *Information integration for concurrent engineering (IICE): IDEF3 Process Description Capture Method Report*. Human Resources Directorate Logistics Research Division, Texas A&M University.

McKenna, R. (1997). *Real time: Preparing for the age of the never satisfied customer*. Boston, MA: Harvard Business School Press.

Melnyk, S. A., Bititci, U., Platts, K., Tobias, J., & Andersen, B. (2015). Is performance measurement and management fit for the future? *Management Accounting Research, 25*(2), 173–186.

Melnyk, S. A., & Christensen, R. T. (2000). *Back to basics: Your guide to manufacturing excellence*. Boca Raton, FL: St. Lucie Press.

Melnyk, S. A., Stewart, D. M., & Swink, M. (2004). Metrics and performance measurement in operations management: Dealing with the metrics maze. *Journal of Operations Management, 22*, 209–217.

Neely, A. (1998). The performance measurement revolution: Why now and what next? *International Journal of Operations and Production Management, 19*(2), 205–228.

Neely, A. (Ed.). (2002). *Business performance measurement: Theory and practice*. Cambridge: Cambridge University Press.

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 and Production Management, 25*(12), 1264–1277.

Neely, A. (Ed.). (2007). *Business performance measurement: Unifying theory and integrating practice*. Cambridge: Cambridge University Press.

Neely, A., & Bourne, M. (2000). Why measurement initiatives fail. *Measuring Business Excellence, 4*(4), 3–7.

Neely, A., Gregory, M., & Platts, K. (2005). Performance measurement system design: A literature review and research agenda. *International Journal of Operations and Production Management, 25*(12), 1228–1263.

Neely, A., Mills, J., Platts, K., Richards, H., Gregory, M., Bourne, M., & Kennerley, M. (2000). Performance measurement system design: Developing and testing a process-based approach. *International Journal of Operations and Production Management, 20*(10), 1119–1145.

New, C. C., & Szwejczewski, M. (1995). Performance measurement and the focused factory: Empirical evidence. *International Journal of Operations and Production Management, 15*(4), 63–79.

NIST (National Institute of Standard and Technology). (2018). Retrieved September 2018, from http://www.quality.nist.gov; http://www.baldrige.nist.gov.

Perkmann, M., Neely, A., & Walsh, K. (2011). How should firms evaluate success in university–industry alliances? A performance measurement system. *R&D Management, 41*(2), 202–216.

Petersen, J. A., Mcalister, L., Reibstein, D. J., Winer, R. S., Kumar, V., & Atkinson, G. (2009). Choosing the right metrics to maximize profitability and shareholder value. *Journal of Retailing, 85*(1), 95–111.

Pfeffer, J., & Salancik, G. R. (1978). *The external control of organizations: A resource dependence perspective*. New York: Harper & Row.

Rametsteiner, E., Pülzl, H., Alkan-Olsson, J., & Frederiksen, P. (2011). Sustainability indicator development—Science or political negotiation? *Ecological Indicators, 11*(1), 61–70.

Skinner, W. (1974, May–June). The decline, fall, and renewal of manufacturing plants. *Harvard Business Review*.

Smith, D. (2000). *The measurement nightmare: How the theory of constraints can resolve conflicting strategies, policies, and measures*. Boca Raton, FL: St. Lucie Press.

Smith, M., & Bititci, U. S. (2017). Interplay between performance measurement and management, employee engagement and performance. *International Journal of Operations and Production Management, 37*(9), 1207–1228.

Stapleton, D., Hanna, J. B., Yagla, S., Johnson, J., & Markussen, D. (2002). Measuring logistics performance using the strategic profit model. *The International Journal of Logistics Management, 13*(1), 89–107.

Taticchi, P., Balachandran, K., & Tonelli, F. (2012). Performance measurement and management systems: State of the art, guidelines for design and challenges. *Measuring Business Excellence, 16*(2), 41–54.

Taylor, J., & Baines, C. (2012). Performance management in UK universities: Implementing the Balanced Scorecard. *Journal of Higher Education Policy and Management, 34*(2), 111–124.

U.S. Department of Energy – PBM SIG (Performance-Based Management Special Interest Group). (2012). *Performance-based management handbook of techniques and tools–How to measure performance (revised)*. Retrieved September 2018, from http://www.orau.gov/pbm.

Ulrich, K., & Eppinger, S. (2000). *Product design and development*. New York: McGraw-Hill.

UNI 11097. (2003). *Indicatori e quadri di gestione della qualità*. Milano.

Use and Abuse of Indicators

2

Abstract

The goal of the present chapter is to introduce the reader to some criticalities concerned with indicators, which are treated in an organic way in the rest of the book. For the purpose of example, the attention is focused on several popular indicators: the HDI (*Human Development Indicator*), some air-quality indicators (AQI, ATMO and IQA) and the scoring indicators used for the Olympic decathlon. These indicators are analysed individually, emphasizing their qualities and drawbacks. Additionally, this chapter mentions some indicator properties that are formalized in detail in Chaps. 3 and 4.

2.1 Introduction

Chapter 1 showed how indicators can be used in a wide range of practical contexts. When reading any newspaper, it seems that these "magic" numbers influence the destiny of the world: "European countries with deficit/GDP ratio lower than 3% can adopt Euro currency"; "country inflation is running at 2.7%"; "the air-quality indicator value is 6, therefore elderly people and children may be at risk", and so on.

Why are indicators considered so important? Presumably, since they are supposed to (properly!) represent reality, consistently with the concept of *faithfulness of representation*. In addition, the use of indicators is practically inevitable when monitoring complex processes.

Let us consider, for example, the HDI (*Human Development Indicator*) indicator, which was introduced by the United Nations Development Programme (UNDP) to measure the development of world countries (Bouyssou et al. 2000; UNDP 2003; 2018). Is the information provided by this indicator independent on the application context? In other words, can the indicator be influenced by the subjects using it (territory planners, administrators, etc.)?

F. Franceschini et al., *Designing Performance Measurement Systems*, Management for Professionals, https://doi.org/10.1007/978-3-030-01192-5_2

In the 1997 annual report, UNDP (1997) cautiously states that "*. . .the HDI has been used in many countries to rank districts or counties as a guide to identifying those most severely disadvantaged in terms of human development. Several countries, such as Philippines, have used it as a planning tool.* [. . .] *The HDI has been used especially when a researcher wants a composite measure of development. For such uses, other indicators have sometimes been added to the HDI . . .*".

What are the real goals of the HDI? Perhaps, to exclude support to those countries that do not correctly plan development? Or to divide the International Monetary Fund aid among poorer countries? Are we sure that HDI is properly defined (according to its representation target)? Are the HDI results significant?

The rest of the chapter tries to answer the above questions, providing a detailed discussion on different indicators: the HDI, some popular air-quality indicators, and the scoring indicators used for Olympic decathlon (Bouyssou et al. 2000). We will analyse each of them individually, trying to emphasize their qualities/drawbacks.

The goal of the present chapter is to introduce the reader to some criticalities concerned with indicators, which will be treated in an organic way in the rest of the book. Additionally, some indicator properties that are mentioned in this chapter will be illustrated in detail in Chaps. 3 and 4.

2.2 Human Development Indicator (HDI)

The HDI is a measure to summarize human development. Although the construction of this indicator has been significantly revised in the year 2010 (Klugman et al. 2011; UNDP 2018), it is "pedagogically" interesting to analyze the construction of the previous version (i.e., the one in use from 1990 to 2009).

In general, the HDI measures the development performance of a country, considering three basic dimensions of human development (UNDP 2003; 2018)[1]:

- a long and healthy life, as measured by life expectancy at birth (*Life Expectancy Indicator*—LEI);
- knowledge (Educational Attainment Indicator—EAI), as measured by the adult literacy rate (*Adult Literacy Indicator*—ALI), which account for 2/3, and the combined primary, secondary and tertiary gross enrolment ratio (ERI), which account for 1/3 of the total amount;
- a decent standard of living, as measured by GDPI (*Gross Domestic Product per capita Indicator*) given as *Purchasing Power Parity* US$ (PPP US$).

Three (sub-)indicators are associated with the above dimensions; all of them are normalized into the range [0, 1] by applying the so-called *min-max normalization*:

[1]Since 1990, *Human Development Report* (HDR) is the annual publication of the *United Nations Development Programme* (UNDP). The 2003 HDR refers to data collected in 2001.

Table 2.1 Reference values for normalizing the HDI (sub)indicators

(Sub)indicator	Name	Unit of measurement	Upper ref. limit	Lower ref. limit	Weight
Life expectancy at birth	LEI	Years	85	25	1
Adult literacy rate	ALI	%	100	0	2/3
Combined gross enrolment ratio	ERI	%	100	0	1/3
GDP per capita	GDPI	PPP$	40,000	100	1

$$\text{(Sub)indicator} = \frac{\text{Measured value} - \text{Lower ref. limit}}{\text{Upper ref. limit} - \text{Lower ref. limit}}. \tag{2.1}$$

Table 2.1 reports the min/max reference limits for calculating the HDI.

The HDI is then calculated through a (weighted) average of (sub)indicators.

The following subsections exemplify the calculation of the HDI for Albania (data related to 2001).

2.2.1 Life Expectancy Indicator (LEI)

The life expectancy indicator (LEI) measures the relative achievement of a country in life expectancy at birth. For Albania, with a life expectancy of 73.4 years in 2001, the life expectancy indicator is:

$$\text{LEI} = \frac{73.4 - 25.0}{85.0 - 25.0} = 0.807. \tag{2.2}$$

2.2.2 Educational Attainment Indicator (EAI)

The education indicator (EAI) measures a country's relative achievement in both adult literacy and combined primary, secondary and tertiary gross enrolment. First, an indicator for adult literacy (ALI) and one for combined gross enrolment (ERI) are calculated (data are referred to the school year 2000/2001):

$$\text{ALI} = \frac{85.3 - 0.0}{100.0 - 0.0} = 0.853; \tag{2.3}$$

$$\text{ERI} = \frac{69.0 - 0.0}{100.0 - 0.0} = 0.690. \tag{2.4}$$

Then, these two indicators are combined to create the education indicator; adult literacy accounting for 2/3 and the combined gross enrolment accounting for 1/3. For Albania, with an adult literacy rate of 85.3% in 2001 and a combined gross enrolment ratio of 69% in the school year 2000/01, the education indicator was 0.799.

$$\text{EAI} = \frac{2 \cdot \text{ALI} + \text{ERI}}{3} = \frac{2 \cdot 0.853 + 0.69}{3} = 0.799. \tag{2.5}$$

2.2.3 Gross Domestic Product Indicator (GDPI)

The GDPI is calculated using adjusted GDP per capita (PPP US$). Considering the HDI, income represents all the remaining dimensions of human development, except long/healthy life and knowledge. Since the value of one dollar is different for people earning $100 in comparison to those earning $100,000, the income has been adjusted (*marginal utility* concept). The income adjustment function is the logarithm function (UNDP 2003). Figure 2.1 shows the effect of this adjustment: the same increase in the adjusted income—*Log* GDP per-capita—determines a little shift of GDP per-capita when the income is low, and a high shift when the income is high.

GDPI is calculated using the following formula:

$$\text{GDPI} = \frac{Log(\text{GDP per capita}) - Log100}{Log40,000 - Log100}. \tag{2.6}$$

For Albania, with a GDP per-capita of $3680 (PPP US$) in 2001, the GDP indicator is (UNDP 2003):

$$\text{GDPI} = \frac{Log3680 - Log100}{Log40,000 - Log100} = 0.602. \tag{2.7}$$

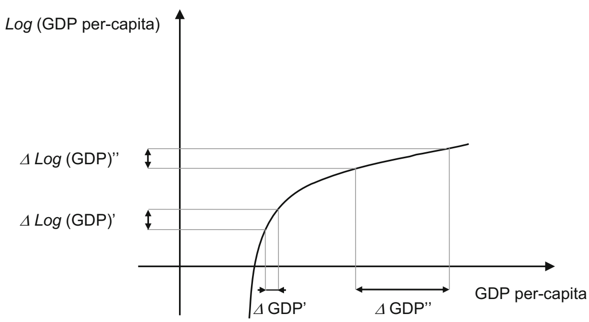

Fig. 2.1 Concept of *marginal utility* of the income per-capita. The same increase in the adjusted income function—*Log*(GDP per-capita)—determines a little shift of GDP per-capita when the income is low and a high shift when the income is high

2.2.4 Calculating the HDI

Once the (sub)indicators related to the three dimensions of interest have been determined, calculating the HDI is straightforward; it is given by the (weighted) average of these sub-indicators (cf. Table 2.1):

$$\mathrm{HDI} = \frac{\mathrm{LEI} + \mathrm{EAI} + \mathrm{GDPI}}{3} = \frac{\mathrm{LEI} + \left(\frac{2 \cdot \mathrm{ALI} + \mathrm{ERI}}{3}\right) + \mathrm{GDPI}}{3} \tag{2.8}$$

E.g., for Albania, the HDI is:

$$\mathrm{HDI} = \frac{0.807 + 0.798 + 0.602}{3} = 0.736 \tag{2.9}$$

2.2.5 Remarks on the Properties of HDI

Scale Normalization

To calculate HDI, the performance in each underlying (sub)indicator (LEI, ALI, ERI and GDPI) is normalized within the interval [0, 1]. Lower and upper reference limits related to each dimension (Table 2.1) are quite arbitrary. Why are life-expectancy limits set to 25 and 85 years? Is 25 years the minimum value registered? Actually, the lowest value ever registered is 22.6 for Rwanda (UNDP 1997); the LEI value for this country would therefore be negative. The lower limit was set to 25 years, at the time of the first UNDP Report (1990), when the lowest value registered was 35. Probably, in that period nobody imagined that the expectancy value could fall below the 25-year limit. To overcome (at least partly) this problem, the limit could be set to a smaller value (for example 20 years).

It is interesting to notice that the choice of (min-max) reference limits has a direct consequence on the HDI calculation. E.g., let us consider Table 2.2, which reports the values of LEI, EAI and GDPI for Turkmenistan and Georgia (UNDP 2003). When the LEI minimum and maximum limits are respectively set to 25 and 85, the resulting HDI is 0.748 for Turkmenistan and 0.746 for Georgia. Reducing the maximum limit to 80, HDI is respectively 0.769 for Turkmenistan and 0.770 for Georgia. This simple adjustment reverses the two countries' HDIs.

Additionally, reducing the life-expectancy reference limits from [25, 85] to [25, 80] makes the LEI values increase with a growth factor of (85–25)/(80–25) $\approx$ 109%. As a consequence, the influence of LEI grows at the expenses of that of EAI and

Table 2.2 Life expectancy, EAI and GDPI for Turkmenistan and Georgia (UNDP 2003)

Country	LEI	EAI	GDPI
Turkmenistan	66.6	0.92	0.63
Georgia	73.4	0.89	0.54

Table 2.3 Values of LEI, ALI, ERI and GDP per-capita for Peru and Lebanon (UNDP 2003)

Country	LEI	ALI	ERI	GDP per-capita
Peru	69.4	0.902	0.83	4570
Lebanon	73.3	0.865	0.76	4170

GDPI. Since the Georgia's LEI value is greater than Turkmenistan's, their HDI ordinal relation is reversed.

Similarly, ALI and ERI limits—which are respectively set to 0 and 100—are arbitrary too, as these values are not likely to be observed in a reasonably recent future. So, the *real* interval is tighter than [0, 100] and the scale values could probably be normalized using more appropriate reference values.

The Effects of Compensation

Let us consider Table 2.3, which reports the values of LEI, ALI, ERI and GDP per-capita for Peru and Lebanon (UNDP 2003). Peru's indicators are greater than Lebanon's, except for LEI. In our opinion, this is a clear sign of underdevelopment of Peru, even if the other indicators are rather good. However, the value of HDI is the same for both countries (0.752). This result is due to the effect of compensation among HDI (sub)indicators: for these two countries, weaker (sub)indicators are compensated by stronger ones, so that the resulting HDI values are identical.

The compensation effect is not reasonable when extremely weak (sub)indicators are compensated by extremely excellent ones. To what extent is such a compensation correct?

Considering LEI and GDP per-capita, a one-year decrease of life expectancy can be compensated by an increase of GDP per-capita. Precisely, a one-year life expectancy decrease is:

$$\Delta(\text{LEI}) = \frac{1}{85 - 25} = 0.0167. \tag{2.10}$$

This variation can be compensated by an increase of GDP per-capita (X), corresponding to:

$$\Delta(\text{LEI}) = \Delta(\text{GDPI}) = \Delta\left(\frac{Log\text{X} - Log100}{Log\,40{,}000 - Log100}\right) = 0.0167, \tag{2.11}$$

from which we obtain:

$$\left(\frac{LogX - Log100}{Log\,40{,}000 - Log100}\right) - \left(\frac{LogX' - Log100}{Log\,40{,}000 - Log100}\right) = 0.0167 \tag{2.12}$$

or

$$\frac{1}{2.602} \cdot Log\frac{X}{X'} = 0.0167. \tag{2.13}$$

The final expression is given by:

$$\frac{X}{X'} = 10^{0.04345} = 1.105. \tag{2.14}$$

This increase of GDP per-capita rate compensates a one-year decrease of life expectancy. For example, when the reference income (X') is \$100, a one-year decrease in life expectancy can be counterbalanced by a higher income $X'' = \$110.5$.

In general, a life-expectancy decrease of n years can be compensated by the following increase in GDP per-capita:

$$X = X' \cdot 10^{n \cdot 0.04345}. \tag{2.15}$$

The term $10^{n \cdot 0.04345}$ is the "substitution rate" between a life-expectancy decrease of n years and a corresponding increase in GDP per-capita.

It can be noticed that the increase in GDP per-capita depends on the reference income (X'):

$$\Delta X = X - X' = X' \cdot \left(10^{n \cdot 0.04345} - 1\right). \tag{2.16}$$

E.g., the increase in GDP per-capita that compensates for a one-year decrease in LEI ($n = 1$) will be $\Delta X = 0.105 \cdot X'$. For example, Congo GDP per-capita is \$970 (UNDP 2003), then:

$$\Delta X = X - X' = 0.105 \cdot X' = \$101.85. \tag{2.17}$$

Spain GDP per-capita is \$20,150 (UNDP 2003), then $\Delta X = \$2115.75$. It seems that the life expectancy of poorer countries (like Congo) is easier to compensate through GDP than that of richer countries (like Spain). Extending this reasoning to the limit, life expectancy of richer countries seems to be worth more than that of poorer countries(!).

Other substitution rates can be derived from Eq. (2.8). For example the relation between LEI and ALI:

$$\Delta(\text{LEI}) = -\frac{2}{3} \cdot \Delta(\text{ALI}). \tag{2.18}$$

In general, a one-year decrease of life expectancy can be compensated by an ALI increase, corresponding to:

$$|\Delta(\text{ALI})| = \frac{3}{2} \cdot |\Delta(\text{LEI})| = \frac{3}{2} \cdot 0.0167 = 0.025 \tag{2.19}$$

Similarly, to compensate a n-year life expectancy decrease, the adult literacy indicator (ALI) has to increase by a $n \cdot 0.025$ factor.

Further Remarks on the GDPI

Focusing on the GDPI model in Eq. (2.6), it can be noticed that GDP per-capita is adjusted using a logarithmic operator. In the past, the GDPI calculation was performed using the so-called Atkinsons' algorithm, as follows (Atkinsons 1970):

$$W(y) = \begin{cases} y & \text{if } 0 < y < y^*, \\ y^* + 2\left[(y - y*)^{1/2}\right] & \text{if } y^* \le y < 2y^*, \\ y^* + 2(y^*)^{1/2} + 3\left[(y - 2y^*)^{1/3}\right] & \text{if } 2y^* \le y < 3y^*, \\ \cdots & \\ y^* + 2(y^*)^{1/2} + 3(y^*)^{1/3} + \cdots + n\left[(y - (n-1)y^*)^{1/n}\right] & \text{if } (n-1)y^* \le y < ny^* \end{cases}, \tag{2.20}$$

being:

y income per-capita;
$W(y)$ transformed income;
y^* annual world average income [for example, in 1994 y^* = \$5835 (UNDP 1997)], depending on the year of interest.

The GDPI value was calculated using the following formula:

$$\text{GDPI} = \frac{W(\text{GDP per capita}) - W(100)}{W(40{,}000) - W(100)}. \tag{2.21}$$

With reference to the year 1994, $W(40{,}000) = 6154$ and $W(100) = 100$.

Comparing the results obtained through the two alternative models [in Eqs. (2.6) and (2.21)], significant differences can be appreciated. For example, in 1994 Greece obtained a \$11,265 income per-capita; using the Atkinsons' algorithm, we obtain:

$$\text{GDPI} = \frac{5982 - 100}{6154 - 100} = 0.972, \tag{2.22}$$

in which the value of 5982 is obtained through the expression:

$$5982 = 5835 + 2 \cdot \left[(11{,}265 - 5835)^{1/2}\right] \quad \text{[see second line of Eq.(2.20)]}. \tag{2.23}$$

On the other hand, when using Eq. (2.6) we obtain:

$$\text{GDPI} = \frac{Log\,11{,}265 - Log\,100}{Log\,40{,}000 - Log\,100} = 0.789. \tag{2.24}$$

Algorithms like the Atkinsons' one [Eq. (2.20)] depend on parameters which may change year after year (value of y^*). As a consequence, comparability among different years is quite difficult. On the other hand, the method in Eq. (2.6) is not affected by this problem.

It is important to remark that, even though these alternative algorithms are (at least apparently) reasonable from an economic point of view, they may produce different results in terms of human development.

Other adjusting functions could have been used to adjust the life expectancy value, the adult literacy rate, and so on. For example, a one-year increase of life expectancy is more relevant for a country where LEI = 30 years than for a country where LEI is 70 years (concept of *marginal utility* for life expectancy).

Statistical Remarks on HDI

Life expectancy is estimated through the mean value of the age of a population; it is therefore not affected by the "shape" of the distribution. For example, in a country where the totality of the population lives up to 50 years, life expectancy is 50 years. In a country where half of the population lives up to 70 years and half up to 30 years, life expectancy is 50 years too. However, the two conditions are very different (Bouyssou et al. 2000).

The above reflection can be extended to the other HDI (sub)indicators. Furthermore, other questions can be raised: What is their actual meaning? Are they all mean values? What is the rationale behind the fusion of (sub)indicators? These questions can be generally referred to all the indicators that fuse different dimensions of a complex process (derived indicators).

2.3 Air Quality Indicators

This section analyses and compares three existing indicators of air quality: the American AQI, the French ATMO, and the Italian IQA.

As a consequence of the dramatic increase in air pollution, especially in large urban areas, many international organizations and governments have adopted multiple regulations to keep the level of pollutants low. Since people incessantly breathe in the atmosphere, the presence of pollutants can be very dangerous for human health.

For several years, researchers have been studying the effect of air pollutants on human health. This activity is carried out monitoring the environment and also using biological indicators to evaluate the impact of pollutants on the population/ecosystem (Rapport et al. 2003; Kampa and Castanas 2008). For instance, several studies established that nitrogen dioxide (NO_2) and ozone (O_3) increase the risk of death in patients with severe asthma (Sunyer et al. 2002; Anderson et al. 2012). Ozone increases the risk of lung cancer (Yang et al. 2005; Lave and Seskin 2013).

Traffic-related air pollution increases mortality (Hoek et al. 2003). Ozone and carbon monoxide (CO) are linked to cardiac birth deficiencies (Ritz et al. 2002; Tankersley et al. 2012), etc.

Public awareness of this problem has gradually increased in the past years due to the media. A high level of air pollution is not only harmful to the population but is also a heavy drain on the wealth of a country. The health damages generate several additional charges, e.g., for health service, mobility and absence from school or work due to sickness, monitoring and protection of the environment, etc.

For this purpose, several countries have introduced evaluation methods that rapidly and efficiently indicate the air quality condition for the population. First, the United States *Environmental Protection Agency* (EPA) developed the Air Quality Indicator—AQI (U.S. EPA 1999; 2016). This indicator provides air quality information using numeric and chromatic indicators, which allow an immediate and extensive evaluation of the risk for the human health.

Other European countries have followed the way. The following sections illustrate the ATMO indicator—developed and implemented by the French Environment Ministry (Ministère de l'écologie et du développement durable 2004)—and the IQA indicator—developed and implemented in some northern regions of Italy, like Piedmont, Lombardy, etc. (Piedmont Regional Law 43/2000 2000). Although these indicators have similar patterns, they also have significant differences. In a nutshell, they take into account the concentrations of the main air pollutants (usually measured in μg/m^3) and set a tolerability scale for each pollutant, trying to gather general information on the overall air condition.

2.3.1 The American Air Quality Indicator (AQI)

The AQI is used for the Metropolitan Statistical Areas (MSAs) of US, with a population of more than 350,000—according to the *Clean Air Act* safety limits of five air pollutants: Ozone (O_3), Particulate Matter (PM), Carbon Monoxide (CO), Sulphur Dioxide (SO_2), Nitrogen Dioxide (NO_2) (U.S. EPA 1999; 2006; 2016).

In each Area, the previous 24-hour concentration of the five mentioned pollutants are measured (or estimated) and reported on six reference categories (Table 2.4).

According to the *Clean Air Act*, an AQI value of 100 or less represents an acceptable concentration for each pollutant. Therefore, AQI values lower than 100 are judged admissible. A higher AQI value means that the air is unhealthy for the more sensitive subjects; as the air pollution increases it also becomes unhealthy for the rest of population.

It is interesting to notice that the above-mentioned thresholds are considerably higher than the EU (European Union) regulations. The EU limits are respectively 120 μg/m^3 for O_3, 150 μg/m^3 for PM_{10} (24-hour mean value), and 100 μg/m^3 for NO_2 (yearly mean value) (European Community, Dir. 96/62/EC, 2002/3/EC).

For each area, the daily AQI value is related to the (sub)indicator concerned with the most critical pollutant:

Table 2.4 Six reference categories of the five pollutants embraced by the AQI (Air Quality Indicator)

O_3	PM_{10}	CO	SO_2	NO_2	
8-hour mean value (μg/m^3)	24-hour mean value (μg/m^3)	8-hour mean value (μg/m^3)	24-hour mean value (μg/m^3)	1-hour mean value (μg/m^3)	AQI reference values
0–137	0–54	0–5.5	0–97	(*)	0–50
138–180	55–154	5.6–11.76	98–412	(*)	51–100
181–223	155–254	11.77–15.5	413–640	(*)	101–150
224–266	255–354	15.6–19.25	641–869	(*)	151–200
267–800	355–424	19.26–38.0	870–1727	1330–2542	201–300
>800	425–604	38.1–50.5	1728–2300	2543–4182	301–500

Calculation method and breakpoints of reference categories are specified for every pollutant (U.S. EPA 1999; 2006; 2016)

(*) U.S. regulations do not set a Nitrogen Dioxide (NO_2) short-term limit. They only define a yearly mean value of 100 μg/m^3. In the AQI calculation the Nitrogen Dioxide is considered uniquely if the hourly mean concentration is higher than 1330 μg/m^3 (so according to AQI reference values upper than 200)

$$\mathrm{AQI} = \max\{I_{O_3}, I_{PM_{10}}, I_{CO}, I_{SO_2}, I_{NO_2}\}. \tag{2.25}$$

The higher the AQI value, the higher the health risk.

Given the pollutant-concentration data and the breakpoints in Table 2.4, every AQI sub-indicator is calculated using Eq. (2.26) (linear interpolation):

$$I_p = I_{L,p} + \frac{I_{H,p} - I_{L,p}}{BP_{H,p} - BP_{L,p}}\left(C_p - BP_{L,p}\right), \tag{2.26}$$

being:

I_p	the sub-indicator for the pth pollutant;
C_p	the concentration of the pth pollutant;
$BP_{H,p}$	the breakpoint that is greater than C_p;
$BP_{L,p}$	the breakpoint that is lower than or equal to C_p;
$I_{H,p}$	the AQI value corresponding to $BP_{H,p}$;
$I_{L,p}$	the AQI value corresponding to $BP_{L,p}$.

For instance, suppose the 8-hour ozone (O_3) concentration is 187 μg/m^3. Then, according to Table 2.6 (see first column and third row), the range that contains this concentration is placed between the breakdowns $BP_{L,O_3} = 181\,\mu g/\mathrm{m}^3$ and $BP_{H,O_3} = 223\,\mu g/\mathrm{m}^3$, and corresponds to the sub-indicator values of $I_{L,O_3} = 101$ to $I_{H,O_3} = 150$. So an ozone concentration of 187 μg/m^3 will correspond to a sub-indicator:

Table 2.5 Air pollutant values registered in a particular metropolitan area

Registered values			
PM_{10} (μg/m^3)	O_3 (μg/m^3)	CO (μg/m^3)	SO_2 (μg/m^3)
158	165	10.5	66

$$\begin{aligned} I_{O_3} &= I_{L,O_3} + \frac{I_{H,O_3} - I_{L,O_3}}{BP_{H,O_3} - BP_{L,O_3}}(C_{O_3} - BP_{L,O_3}) \\ &= 101 + \frac{150 - 101}{223 - 181}(187 - 181) = 108. \end{aligned} \tag{2.27}$$

Let us now consider the air condition in Table 2.5.

$$\begin{aligned} I_{PM_{10}} &= I_{L,PM_{10}} + \frac{I_{H,PM_{10}} - I_{L,PM_{10}}}{BP_{H,PM_{10}} - BP_{L,PM_{10}}}(C_{PM_{10}} - BP_{L,PM_{10}}) \\ &= 101 + \frac{150 - 101}{254 - 155}(158 - 155) = 102 \end{aligned} \tag{2.28}$$

$$\begin{aligned} I_{O_3} &= I_{L,O_3} + \frac{I_{H,O_3} - I_{L,O_3}}{BP_{H,O_3} - BP_{L,O_3}}(C_{O_3} - BP_{L,O_3}) \\ &= 51 + \frac{100 - 51}{180 - 138}(165 - 138) = 82 \end{aligned} \tag{2.29}$$

$$\begin{aligned} I_{CO} &= I_{L,CO} + \frac{I_{H,CO} - I_{L,CO}}{BP_{H,CO} - BP_{L,CO}}(C_{CO} - BP_{L,CO}) \\ &= 51 + \frac{100 - 51}{11.76 - 5.6}(10.5 - 5.6) = 90 \end{aligned} \tag{2.30}$$

$$\begin{aligned} I_{SO_2} &= I_{L,SO_2} + \frac{I_{H,SO_2} - I_{L,SO_2}}{BP_{H,SO_2} - BP_{L,SO_2}}(C_{SO_2} - BP_{L,SO_2}) \\ &= 0 + \frac{50 - 0}{97 - 0}(66 - 0) = 34 \end{aligned} \tag{2.31}$$

The AQI is 102, PM_{10} being the most critical pollutant:

$$\text{AQI} = \max\{I_{PM_{10}}, I_{O_3}, I_{CO}, I_{SO_2}\} = \max\{102, 82, 90, 34\} = 102. \tag{2.32}$$

Each AQI value is linked with a colour and a descriptor. The AQI scale is split into six reference categories by EPA (cf. Table 2.4). The more the AQI value increases, the more the population health risk increases (see Table 2.6).

For example, when the AQI is 50, the air quality is good with a low risk level, and the associated colour is green. For an AQI higher than 300, the risk level is considerably high and the associated colour is maroon.

The EPA qualitative description related to the AQI categories are:

- "*Good*": the AQI value is within the 0–50 range. The air quality is satisfactory, with very little risk to the population.

Table 2.6 American AQI categories, descriptors, and colours (U.S. EPA 1999, 2006, 2016)

AQI reference values	Descriptor	Colour
0–50	Good	Green
51–100	Moderate	Yellow
101–150	Unhealthy for sensitive groups	Orange
151–200	Unhealthy	Red
201–300	Very unhealthy	Purple
>301	Hazardous	Maroon

- "*Moderate*": AQI included between 51 and 100. The air quality is admissible, however a few people could be damaged because of the presence of pollutants. For instance, ozone sensitive people may experience respiratory symptoms.
- "*Unhealthy for sensitive groups*": children and adults with respiratory disease are at risk when doing outdoor activities, due to the ozone exposure, whereas people with cardiovascular disease are most at risk due to the carbon monoxide exposure. When the AQI value is included between 101 and 150, these sensitive individuals could increase their symptoms of disease to the point of health compromising. However, much of the population is not at risk.
- Within the 151–200 range the AQI is considered to be "*unhealthy*". This situation causes possible disease for the general population. Sensitive individuals could seriously suffer.
- "*Very unhealthy*" AQI values—between 201 and 300—represent an alarm. The whole population could be health damaged seriously.
- "*Hazardous*"—over 300—AQI values trigger an immediate alarm. The whole population is at serious risk of suffering.

This indicator puts the emphasis of the risk on the most sensitive groups, like children, elderly people and people with respiratory or cardiovascular disease. Adequate communications support the AQI utilization and have made it available to population. The colour format—which represents the listed categories, the pollution level and the linked health risk—allows people to react depending on circumstances.

AQI Non-monotony

Referring to the example data in Table 2.5, let us suppose that CO concentration increases, driving the indicator I_{CO} from 30 to 102. This new condition is surely worse than the previous one (two of the four sub-indicators have a value of 102). Nevertheless, the AQI indicator remains unchanged and it does not properly represent any significant air pollution change. Therefore, the AQI does not fulfil the property of monotony. This property will be formalized and illustrated in more detail in Sect. 4.6.3.

AQI Non-compensation

Suppose we were to calculate the AQI taking into account the two different air quality conditions (W and Z) in Table 2.7.

Table 2.7 AQI sub-indicators values in two air-quality conditions

Condition	PM_{10}	O_3	CO	SO_2
(W)	155	30	25	30
(Z)	130	104	100	121

Although some pollutant concentrations (O_3, CO and SO_2) are worse in condition Z, the AQI is higher in condition W

The first set of data (W) is almost perfect except for the PM_{10} concentration, whereas the second one (Z) is not particularly good for all the sub-indicators. Nevertheless, the AQI is 155 in condition (W) and 130 in condition (Z). Unlike other indicators, the AQI does not fulfil any form of compensation.

AQI Scale Levels

Referring to Table 2.6, it is clear that the "mapping" between each pollutant concentration bandwidth and the corresponding sub-indicator value is not homogeneous.

Considering, for example, the SO_2 pollutant, the first AQI level [0–50] is related to the [0–97 μg/m^3] range of concentration and the pollutant, the second AQI level [51–100] is related to the [98–412 μg/m^3] range of concentration of the pollutant. While the ranges of AQI levels are constant, the ranges of the relevant concentrations differ.

The size of the range of each pollutant is supposed to be fixed on the basis of the effects on human health. This assumption partly contrasts with the direct proportionality assumption between the air-pollutant concentration and the AQI within each specific range [see Eq. (2.27)].

2.3.2 The ATMO Indicator

The ATMO indicator, which was developed by the French Environment Ministry, is based on the concentration of four air pollutants: Ozone (O_3), Particulate Matter (PM), Sulphur Dioxide (SO_2), and Nitrogen Dioxide (NO_2) (Ministère de l'écologie et du développement durable 2004; Bouyssou et al. 2000). The concentration of each of the pollutants is related to a sub-indicator that is expressed on a ten level scale. The first level corresponds to an excellent air quality, the fifth and sixth level are just around the European long-term norms, the eighth level corresponds to the EU short-term norms, and the tenth level corresponds to a health hazard condition. The ten-level scales of sub-indicators are shown in Table 2.8.

The ATMO value is the maximum of four sub-indicators:

$$\text{ATMO} = \max\{I_{NO_2}, I_{SO_2}, I_{O_3}, I_{PM_{10}},\}. \tag{2.33}$$

The PM_{10} value is the (grand) average of the daily average values, registered from 1:00 to 24:00 by different stations in the monitored area. The value of the other pollutants is the average of the maximum hourly values, registered from 1:00 to 24:00 by different stations in the monitored area.

Table 2.8 Ten reference categories of the four sub-indicators contemplated by the ATMO indicator

Level	PM_{10} (μg/m^3)		O_2 (μg/m^3)		O_3 (μg/m^3)		SO_2 (μg/m^3)		Descriptor	Colour
	Min	Max	Min	Max	Min	Max	Min	Max		
1	0	9	0	29	0	29	0	39	Very Good	Green
2	10	19	30	54	30	54	40	79	Very Good	Green
3	20	29	55	84	55	79	80	119	Good	Green
4	30	39	85	109	80	104	120	159	Good	Green
5	40	49	110	134	105	129	160	199	Medium	Orange
6	50	64	135	164	130	149	200	249	Poor	Orange
7	65	79	165	199	150	179	250	299	Poor	Orange
8	80	99	200	274	180	209	300	399	Bad	Red
9	100	124	275	399	210	239	400	499	Bad	Red
10	≥125		≥400		≥240		≥500		Very Bad	Red

Each level is included between a minimum and a maximum breakpoint value (Ministère de l'écologie et du développement durable 2004)

Table 2.9 Sub-indicators determining a resulting ATMO value of 8

Pollutant	NO_2	SO_2	O_3	PM_{10}
Sub-indicator	2	2	3	8

Table 2.10 ATMO sub-indicator values for two air-quality conditions

Condition	NO_2	SO_2	O_3	PM_{10}
(U)	2	1	7	1
(V)	6	5	6	6

Although in condition Y some pollutant concentrations (NO_2, SO_2 and PM_{10}) are worse than in condition X, the ATMO indicator is higher in the former condition

For the purpose of example, consider the air condition in Table 2.9. Consistently with the model in Eq. (2.33), the ATMO value is 8 and therefore the air quality is very unhealthy.

A brief description of the ATMO main properties is reported in the following subsections.

ATMO Non-monotony

Let imagine a sunny day with heavy traffic and no wind. Referring to the example data in Table 2.9, let us suppose the Ozone concentration increases, driving the corresponding indicator from 3 to 8. The new condition is surely worse than the previous one (two of the four sub-indicators have a value of 8). Even though the ATMO value would probably be supposed to increase, it remains unchanged, without representing any significant air-pollution variation. This proves that the ATMO does not fulfil the property of (strict) monotony. This property will be formalized in Sect. 4.6.3.

ATMO Non-compensation

Let us exemplify the calculation of the ATMO in the two different air-quality conditions (U and V) in Table 2.10.

Data in condition (U) are almost perfect except for the O_3 concentration, whereas in condition (V) they are not particularly good for all (sub)indicators. Nevertheless, the ATMO is 7 in condition (U) and 6 in condition (V).

In this case, the ATMO (sub)indicators do not compensate each other. In the state U, the relatively high O_3 value is not counterbalanced by the low values of the three remaining (sub)indicators.

ATMO Scale Levels

Likewise AQI, the ATMO indicator is characterized by a non-homogeneous "mapping" of the concentration of each pollutant into a corresponding (sub)indicator value (see Table 2.8). For example, considering SO_2, the first ATMO level is related to the [0–39 $\mu g/m^3$] range of concentration of the pollutant (width of 39 $\mu g/m^3$), while the sixth ATMO level is related to the [200–249 $\mu g/m^3$] range of concentration (width of 49 μg/m^3). Since the width of the range of pollutants is supposed to be determined on the basis of the negative effects on human health, pollutant concentrations in the same ATMO level are supposed to be equivalent; e.g., an SO_2 concentration of 201 μg/m^3 has to be considered equivalent to a 249 μg/m^3 one.

Additionally, it can be seen that the concentration reference values of ATMO are considerably smaller than the AQI ones.

2.3.3 The IQA Indicator

In some northern Italian regions (Piedmont, Lombardy, etc.) different protocols are currently being tested to monitor air quality and inform the population. We now focus on the IQA indicator (Indice di Qualità dell'Aria), which has been used in Piedmont (Piedmont Regional Law 43/2000 2000). This indicator is inspired by the AQI (in Sect. 2.3.1) but has some clear differences.

According to the safety regulation limit, the IQA aggregates the most critical air pollutants on the basis of their effects on human health: Ozone (O_3) and Particulate Matter (PM_{10}) in summertime, PM_{10} and Nitrogen Dioxide (NO_2) in wintertime.

The IQA can be associated with a level between 1 and 7; the higher the air pollution, the higher the health hazard and the relevant indicator level. Consequently, IQA can also be considered as an indicator of human-health risk.

The IQA includes several sub-indicators, which are related to the different pollutants of interest: Ozone (O_3), Particulate Matter (PM_{10}) and Nitrogen Dioxide (NO_2). IQA is defined as the arithmetic mean of the two highest sub-indicators:

$$\mathrm{IQA} = \frac{I_1 + I_2}{2}, \tag{2.34}$$

I_1 and I_2 being the two highest sub-indicators.

The IQA is calculated on a daily basis using the concentration of pollutants in the previous 24 hours. The resulting IQA value is then associated with a 7-level scale, as shown in Table 2.11. This final level is presented to the population; this information is enriched by an indication of the air pollution evolution, resulting from weather forecast.

Summarizing, IQA is a conventional indicator used to:

- report the quality of the air on a daily basis;
- identify the worst environmental parameters;
- estimate the risk for the population.

An IQA of 100 represents a threshold value for human-health risk. IQA values lower than 100 are generally satisfactory with no potential hazard for human health. The more the IQA value exceeds 100, the more the air quality is considered unhealthy (initially for the most sensitive groups only and then for the rest of the population).

Table 2.11 IQA categories, descriptors and colours (Piedmont Regional Law 43/2000 2000)

IQA reference values	IQA final level	Descriptor	Colour
0–50	1	Excellent	Blue
51–75	2	Good	Light blue
76–100	3	Fair	Green
101–125	4	Mediocre	Yellow
126–150	5	Not very healthy	Orange
151–175	6	Unhealthy	Red
>175	7	Very unhealthy	Purple

Different descriptions of air quality, different colours, and useful advices for the population are generally associated with each of the seven IQA levels:

- "*Excellent*" (blue), with a reference value between 0 and 50. The quality of the air is considered excellent.
- "*Good*" (light blue), with a reference value between 51 and 75. The air quality is considered very satisfactory with no risk for the population.
- "*Fair*" (green), with a reference value between 76 and 100. The air quality is satisfactory and there is no risk for the population.
- "*Mediocre*" (yellow), with a reference value between 101 and 125. The population is not at risk. However, people with asthma, chronic bronchitis or heart problems might show some symptoms during intense physical activity; it is advised that these people limit their physical exercise outdoors, especially in summertime.
- "*Not very healthy*" (orange), with a reference value between 126 and 150. People with heart problems, elderly people and children may be at risk. These categories of people should limit their physical activity outdoors, especially at peak times in summer.
- "*Unhealthy*" (red), with a reference value between 151 and 175. Many people could have slightly negative health problems, albeit reversible; it is advised to limit extended periods of time outdoors, especially at peak times in summer. Unfortunately, people in the sensitive groups could have more serious symptoms; in these cases it is highly recommended to limit outdoor activity as much as possible.
- "*Very unhealthy*" (purple), with a reference value above 175. There may be slightly negative effects on the health of the whole population. Elderly people and people with breathing difficulties should avoid going outside. Other people (especially children) should avoid physical activity outdoors, especially at peak times in summer.

The IQA sub-indicators are:

- *Nitrogen dioxide* (NO_2) sub-indicator:

$$I_{NO_2} = \frac{\overline{V}_{\max h_{NO_2}}}{V_{\text{ref } h_{NO_2}}} \cdot 100. \tag{2.35}$$

$\overline{V}_{\max h_{NO_2}}$ is the average of the maximum hourly NO_2 concentrations registered from 1:00 to 24:00 by different stations in the monitored area;

$V_{\text{ref } h_{NO_2}}$ is a NO_2 reference hourly concentration of 200 μg/m^3, which represents a safety limit (Ministero dell'ambiente e della tutela del territorio 2002).

- *Particulate matter* (PM_{10}) sub-indicator:

$$I_{PM_{10}} = \frac{\overline{V}_{\text{avg } 24h_{PM_{10}}}}{V_{\text{ref}_{PM_{10}}}} \cdot 100. \tag{2.36}$$

$\overline{V}_{\text{avg } 24h_{PM_{10}}}$ is the (grand) average of the average hourly PM_{10} concentrations, registered from 1:00 to 24:00 by different stations in the monitored area;

$V_{\text{ref}_{PM_{10}}}$ is a PM_{10} reference daily concentration of 50 μg/m^3, which represents a safety limit (Ministero dell'ambiente e della tutela del territorio 2002).

- *Ozone* (O_3) sub-indicator:

$$I_{8hO_3} = \frac{\overline{V}_{\max 8h_{O_3}}}{V_{\text{ref } 8h_{O_3}}} \cdot 100. \tag{2.37}$$

$\overline{V}_{\max 8h_{O_3}}$ is the average of the maximum O_3 concentrations, calculated every hour on the basis of the data collected in the previous 8 hours (rolling sample) by different stations in the monitored area;

$V_{\text{ref } 8h_{O_3}}$ is a O_3 reference concentration of 120 μg/m^3, which represents a safety limit (Dir. 2002/3/EC).

To assess the evolution of the atmospheric pollution, the IQA level registered in a certain day is associated with the IQA levels registered in the six previous days.

For the purpose of example, let show the calculation of the IQA indicator for the Turin's metropolitan area, in 27 January 2005. Data related to each pollutant are reported in Table 2.12 (Province of Turin's Regional Agency for the Environment—ARPA 2005).

The first three pollutants are used to calculate the IQA value. The relevant sub-indicators are respectively:

Table 2.12 Air-pollutant values registered in the Turin's metropolitan area on 27 January 2005 (ARPA, Province of Turin 2005)

PM_{10} (μg/m^3)	O_3 (μg/m^3)	NO_2 (μg/m^3)	C_6O_6 (μg/m^3)	CO (μg/m^3)	SO_2 (μg/m^3)
84	33	125	5.6	2.2	15

$$\begin{aligned} I_{NO_2} &= \frac{125}{200}\cdot 100 = 62.5; \quad I_{PM_{10}} = \frac{84}{50}\cdot 100 = 168; \\ I_{8hO_3} &= \frac{33}{120}\cdot 100 = 27.5 \end{aligned} \tag{2.38}$$

It can be noticed that the two highest sub-indicators are PM_{10} and NO_2. The IQA indicator value is:

$$\text{IQA} = \frac{I_{NO_2} + I_{PM_{10}}}{2} = \frac{62.5 + 168}{2} = 115.3. \tag{2.39}$$

This value corresponds to level 4 (mediocre), according to the IQA scale in Table 2.11.

IQA Non-monotony

The IQA indicator is not (strictly) monotonic with respect to sub-indicators. Considering the data in Table 2.12, let us suppose that the sub-indicator related to O_3 goes from 33 to 61. This new condition (which is obviously worse than the previous one) is still described by the same IQA value (115.3). In this case, the IQA does not "respond" to significant changes in the concentration of pollutants.

IQA Compensation

Let us calculate the IQA considering the two different air quality conditions (*X* and *Y*) in Table 2.13. The relevant sub-indicators are respectively:

$$\begin{aligned} I_{NO_2} &= \frac{125}{200}\cdot 100 = 62.5; \quad I_{PM_{10}} = \frac{84}{50}\cdot 100 = 168; \\ I_{8hO_3} &= \frac{33}{120}\cdot 100 = 27.5 \end{aligned} \tag{2.40}$$

$$\begin{aligned} I_{NO_2} &= \frac{336}{200}\cdot 100 = 168; \quad I_{PM_{10}} = \frac{31.25}{50}\cdot 100 = 62.5; \\ I_{8hO_3} &= \frac{33}{120}\cdot 100 = 27.5 \end{aligned} \tag{2.41}$$

In both conditions, the two highest sub-indicators are PM_{10} and NO_2, which drive the IQA to the same value:

Table 2.13 Concentrations of pollutants in two conditions. Although the concentrations are different, the IQA value is the same

Registered values			
Condition	PM_{10} (μg/m^3)	O_3 (μg/m^3)	NO_2 (μg/m^3)
(X)	84	33	125
(Y)	31.25	33	336

$$\text{IQA} = \frac{I_{NO_2} + I_{PM_{10}}}{2} = \frac{62.5 + 168}{2} = 115.3. \tag{2.42}$$

Based on the health risk estimated by the IQA indicator, the two previous conditions are considered equivalent. We can therefore define a substitution rate between sub-indicators [cf. Eq. (2.34)]:

$$\Delta(I_{NO_2}) = -\Delta(I_{PM_{10}}). \tag{2.43}$$

Combining Eqs. (2.35) and (2.36), we obtain:

$$\frac{\Delta C_{NO_2} \cdot 100}{200} = -\frac{\Delta C_{PM_{10}} \cdot 100}{50}, \tag{2.44}$$

being:
ΔC_{NO_2} variation in the NO_2 concentration,
$\Delta C_{PM_{10}}$ variation in the PM_{10} concentration,

from which:

$$\Delta C_{NO_2} = -4 \cdot \Delta C_{PM_{10}}. \tag{2.45}$$

I.e., a 1 μg/m^3 variation in the concentration of PM_{10} is balanced by a 4 μg/m^3 variation in the concentration of NO_2. Considering the damages to human health, is this substitution rate reasonable?

IQA Scale Levels

According to the formulae in Eqs. (2.35), (2.36) and (2.37), the sub-indicators of IQA are directly proportional to the relevant pollutant concentrations.

However, the calculation of each sub-indicator is influenced by the reference values established by the regulations for the protection of human health (e.g., D.M 2.04.2002 n. 60 and Dir 2000/3/EC). Possible changes of the reference values may have direct consequences on the sub-indicators and their comparability over time.

Let us focus on Eq. (2.34), which expresses the IQA indicator as with the average of the two most critical sub-indicators. What are the scale properties of the sub-indicators of IQA? Being obtained through the arithmetic average of two sub-indicators, the conventional IQA scale is supposed to have the *interval* property (cf. Sect. 3.2.3). Is this right? Is the proposed aggregation mechanism really the best way of modelling the effects of pollution on human health?

Finally, we notice that the reference values of PM_{10}, O_3 and NO_2 for the calculation of IQA are considerably lower than those for AQI.

2.3.4 Comments on the Meaning of (sub)Indicators

Let us consider the statement: "Today's AQI value is twice (150) as much as yesterday's (75)". Does it make sense? Which typology of scaling does the AQI adopt?

Let us go back to the definition of AQI sub-indicators. The conversion of the concentrations of pollutants into the AQI scale (Table 2.4) is arbitrary. For instance, instead of converting the O_3 concentrations of [0–137 μg/m^3] into [0–50] and the concentrations of [138–180 μg/m^3] into [51–100], they could be conventionally converted to [0–40] and [41–80] respectively.

In general, air-quality indicators are supposed to depict the potential (negative) effects of air pollution on human health. The seven-level conversion is only a way to make the information practical to the population. Since the sub-indicator scales are ordinal, the statement "Today the ozone sub-indicator is higher than yesterday" is reasonable (Franceschini et al. 2005); on the other hand, the statement: "Today the AQI indicator is higher than yesterday" is dubious. Regarding the effects on human health, can an AQI value, which is mainly caused by the relatively high concentration of SO_2, be compared with an identical value, which is mainly caused by the relatively high concentration of PM_{10}? In addition, it can be noticed that the AQI does not fulfil the property of strict monotony (formalized in Sect. 4.6.3).

Some questions arise when different sub-indicators result into the same value of the aggregated indicator. For example, to what extent is it reasonable to consider the corresponding negative effects on human health as equivalent? Competent authorities have probably built the sub-indicator scales considering the effects of individual pollutants separately. A question on the concept of equivalence is therefore still open.

Is there any interaction among pollutants? Perhaps some pollutants could damage human health in the short term, others in the long term; the effects can be different on the different parts of the body, and so on. How can the possible damages of pollution be estimated? Perhaps considering health-care cost, mortality rate, etc.?

2.3.5 Comparison of Air-Quality Indicators

The previous sections have shown similarities and differences among the three air-quality indicators: AQI, ATMO, IQA. Although the individual pollutants are similar (PM_{10}, O_3, NO_2, SO_2, . . .), the aggregated indicators may differ in terms of:

- Calculation of sub-indicators;
- Number of risk classes;

- Reference categories of pollutant concentrations;
- Aggregation model of sub-indicators;
- Possible compensation between sub-indicators.

These differences prove that the same physical phenomenon can be represented in different ways. The following chapters will provide an organic discussion on the problem of the *uniqueness* of representation.

A common aspect of the afore-discussed aggregated indicators is that they all do not make the predominant pollutant explicit. For example, knowing that ATMO indicator is 6, we do not necessarily have any information on the predominant sub-indicator(s). The same goes for AQI and IQA.

It is worth remarking that indicators are not measurements, although they are defined on the basis of physical values (see Chaps. 3 and 5). Given a particular combination of air pollutants, each of the three indicators leads to a different result, as it maps the pollutant concentrations into a different scale, whose number of levels and range values are absolutely conventional (see Chap. 3).

Another significant aspect of these alternative air-quality indicators is that they may consider different combinations of air pollutants as equally dangerous for human health (see, for example, the section discussing the ATMO non-monotony). In other terms, comparable values of these indicators do not necessarily entail comparable concentrations of the individual air pollutants.

2.4 The Decathlon Competition

Indicators play an important role in sport competitions. In many sports, specific indicators are used to determine the final ranking of single competitions or entire championships. Let us consider, for example, formula-one racing, tennis, alpine skiing racing, artistic gymnastics, synchronized swimming, etc. (Maisano et al. 2016; Lins et al. 2003; Bouyssou et al. 2000). Particularly interesting is the analysis of the scoring method related to Decathlon competition.

Decathlon is an athletic competition containing 10 different athletic contests and the winner is the participant which amasses the highest overall score. In other words, Decathlon is a two-day miniature track meet, designed to ascertain the best all-around athlete. Within its competitive rules, each athlete must sprint for 100 meters, long jump, heave a 16-pound shotput, high jump and run 400 meters—all in that very order—on the first day. On the second day, the athlete runs a 110 meter hurdle race over 42 inch barriers, hurls the discus, pole vaults, tosses a javelin and, at the end of the contest, races over 1500 meters, virtually a mile.

Decathlon was first introduced at the 1912 Olympic Games of Stockholm, as a three-day multi-event contest. At the begin, the score of the single event was given by the order of arrival. For example, if an athlete finished third in a particular event, he gained three points. The winner of the competition was the athlete with the lowest overall score (Zarnowsky 1989). This approach is close to the score aggregation method proposed by Jean-Charles de Borda (see Chap. 3) (Vansnick 1986; Roy 1996). The main drawback of this method is that it overlooks the athletes' performance level. An athlete who arrives 0.1, 1, or 10 seconds before the next one, finishing a contest in *i*-th position, always gains *i* points, and *i* + 1 the next athlete, independently of the performance level.

The problem has been solved introducing some merit scoring tables, connecting the absolute result of each event with a specific score (for instance the time of 10.00 seconds in the 100 metres is actually worth 1096 points). Unlike the first scoring method (based on the order of arrival), in this other case, the higher the score the better the performance. So, the winner of the competition is the athlete who has scored the highest number of points.

The scoring tables have been published in different versions over the years. At first, the construction of tables was based on the idea that when the performance is close to the world record, it is more difficult to improve it. Therefore, as the performance increases, the score increases more than proportionally.

Figure 2.2 shows a graphic representation ("convex" curve) of the scoring tables initially suggested by IAAF (International Association of Athletics Federations). The example refers to *distance*-based events. For *time*-based events, the graph shape is different: as time spent decreases, the score increases proportionally.

But the "convex" scoring tables may generate some problems. If one athlete specializes in a subset of events (for example 4 events) neglecting the others, he may have an advantage over the other athletes. By obtaining results close to the world records in his favourite event(s), he will score many points, intentionally neglecting other events: poor scores in unfavourable contests will be largely compensated by high scores in the favourable ones (see the higher curve slope in Fig. 2.2). Consequently, it is more convenient to specialize in few contents, rather than preparing for them all, with the same commitment. But this is not in the spirit of decathlon, which consists in recompensing eclecticism, i.e., the ability of performing well in all different events.

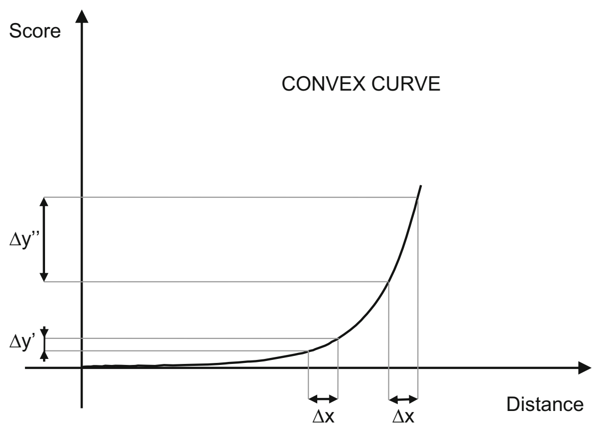

Fig. 2.2 Graphical representation of the decathlon scoring tables (referred to distance-based events) in use in the 1934-to-1962 period. When the performance is close to the world record, the score increases more than directly proportionally (convex curve)

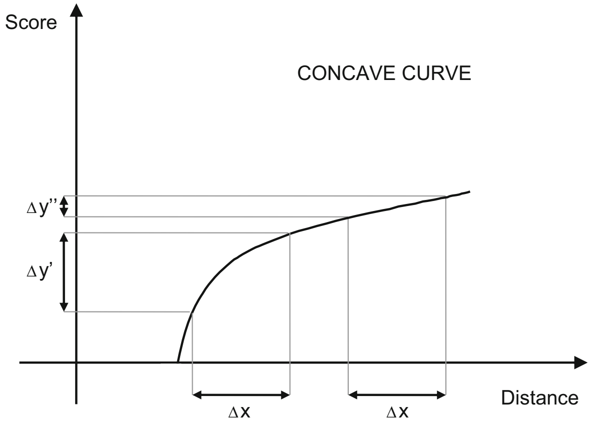

Fig. 2.3 Graphical representation of the decathlon scoring tables (referred to distance-based events) in use after 1962. The purpose of the curve is to discourage poor performances (concave curve)

To prevent this problem, in 1962 the International Olympic Committee suggested new scoring tables, which can be graphically represented as curves with concave profile (Fig. 2.3). The purpose is to discourage poor performances, stimulating athletes' commitment in the totality of the events.

Since 1962, scoring curves have been periodically refined. By way of example, Fig. 2.4a, b, shows the scoring curves for the high jump and the 100 metres, updated on 2017 (IAAF 2017).

The analysis of decathlon scoring tables stimulates the following considerations (Bouyssou et al. 2000):

- How to determine reference values (minimum and maximum) of the scoring tables? As already discussed for HDI, the choice of these values directly impacts on the total score calculation (see Sect. 2.2.5). Furthermore, since maximum values generally depend on the event's world record, the scoring tables will inevitably change over time.
- What is the rationale behind aggregating scores through an additive model? Could a multiplicative aggregation model better encourage good performances in all events? To explain the concept, let us consider a *triathlon* competition (similar to decathlon but with three events only) where each event is associated with a 0-to-10 score. One athlete (X) gets 8 points in all the three events, while another one (Y) gets the scores 9, 8 and 7. If we use an additive rule, both athletes totalize 24 points. If we use a multiplicative rule, X gets 512 points, while Y gets 504 points only. In this case, the athlete with higher performance in all the events is favoured—as in the spirit of decathlon.

2.4.1 The "Normative" Effect of Scoring Indicators

Although the main function of decathlon scoring rules is to determine the winner of a competition, these rules may affect racing strategies.

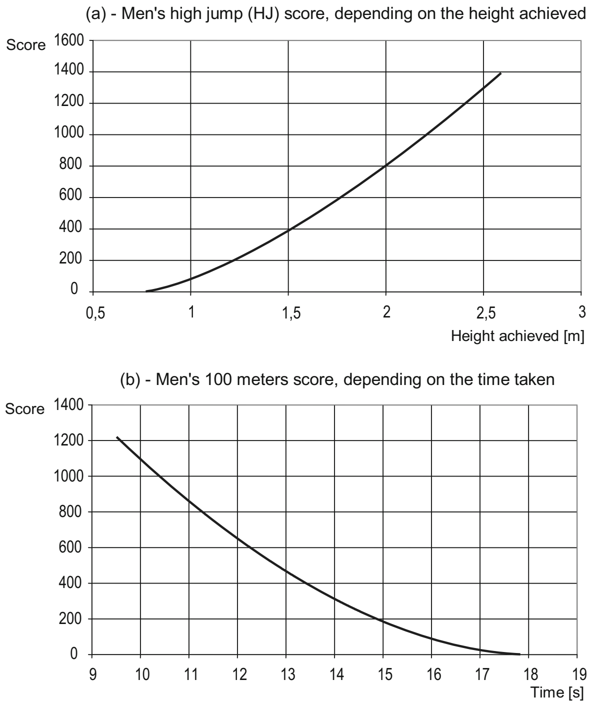

Fig. 2.4 Scoring curves for the high jump (**a**) and the 100 meters (**b**) (IAAF 2017)

First of all, scoring indicators are often used to evaluate the athletes' performance: what matters is not the performance level but the related score. When one athlete gets a score close to the winner's, he is considered a good athlete. This statement is basically true, but it should be analysed in more detail. In fact, the overall score "hides" the individual contributions in single events; e.g., low scores can be obtained by athletes who are outstanding in some events but inadequate in other ones.

A second aspect concerns athletes' preparation. Scores can influence preparation strategies and the definition of medium/long-term objectives. At the beginning, when the final victory was achieved on the basis of single-event rankings, the athletes' main goal was overtaking the others. Each race was a separate matter and it was not unusual to win one event against mediocre athletes, and lose another event against better ones.

A method to estimate the overall performance level was lacking. The scoring tables introduced later have changed the importance and the meaning of single-race results. Athletes not only compete to win races, but also to improve themselves. Apart from designating the winner, scoring tables are useful tools for comparing the results and plan the athletic preparation. In other words, the scoring indicators have gradually become tool for monitoring the preparation of the athletes.

These considerations show that, although indicators are used to describe a process (decathlon competition in this case), they may significantly influence it. In other words, the "representational" role of indicators may often become "normative".

Table 2.14 Decision and representation: analogies and differences

Decision	Criteria	Alternatives
Representation	Indicators	Process conditions

2.4.2 Representation and Decision

In all complex decision-making problems, the main concern of a decision maker is to fulfil his/her conflicting goals while satisfying the possible constraints (Vincke 1992; Roy 1993). The literature offers many different approaches for analysing and solving decision-making problems.

In this context, indicators are used to represent the different conditions of a process; this representation is fundamental for the process analysis and control (management). The success of the representation depends on the choice of indicators and their effectiveness in representing the process of interest.

The analogy between decision-making models and representational models is evident: their common purpose is to classify a set of objects or "alternatives", synthesising multi-dimensional information (see Table 2.14). However, there is a big difference: the construction of indicators usually does not require any decision-problem formalization, decision maker, etc.

The field of action of indicators is larger than that of decision-making models. Indicators are not typically used for decision making but sometimes they can help it.

References

Anderson, J. O., Thundiyil, J. G., & Stolbach, A. (2012). Clearing the air: A review of the effects of particulate matter air pollution on human health. *Journal of Medical Toxicology, 2*, 166–175.

ARPA – Piedmont's Regional Agency for the Environment. (2005). *Tabella dei valori massimi in Torino - gennaio 2005*. ARPA environmental vigilance and control network, Turin, Italy.

Atkinsons, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory, 2*, 244–263.

Bouyssou, D., Marchant, T., Pirlot, M., Perny, P., Tsoukiàs, A., & Vincke, P. (2000). *Evaluation and decision Models: A critical perspectives*. Boston: Kluwer Academic Publishers.

Franceschini, F., Galetto, M., & Maisano, D. (2005). A short survey on air quality indicators: Properties, use, and (mis)use. *Management of Environmental Quality: An International Journal, 16*(5), 490–504.

Hoek, G., Brunekreef, B., Goldbohm, S., Fischer, P., & Van Den Brandt, P. A. (2003). Cardiopulmonary mortality and air pollution. *The Lancet, 362*(9384), 587.

IAAF (International Association of Athletics Federations). (2017). *Scoring tables*. Retrieved September 2018, from www.iaaf.org.

Kampa, M., & Castanas, E. (2008). Human health effects of air pollution. *Environmental Pollution, 151*(2), 362–367.

Klugman, J., Rodríguez, F., & Choi, H. J. (2011). The HDI 2010: New controversies, old critiques. *The Journal of Economic Inequality, 9*(2), 249–288.

Lave, L. B., & Seskin, E. P. (2013). *Air pollution and human health*. New York: RFF Press.

Lins, M. P. E., Gomes, E. G., Soares De Mello, J. C. B., & Soares De Mello, A. J. R. (2003). Olympic ranking based on a zero sum gains DEA model. *European Journal of Operational Research, 148*(2), 85–95.

Maisano, D., Botta, A., & Franceschini, F. (2016). On the rating system in alpine skiing racing: Criticism and new proposals. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology, 230*(4), 253–263.

Ministère de l'écologie et du développement durable. (2004). Arrêté du 22 julliet 2004 relatif aux indices de la qualité de l'air. Journal official de la Rèpublique Française, France.

Ministero dell'ambiente e della tutela del territorio. (2002). Decreto 2 aprile 2002 n. 60, Recepimento della direttiva 1999/30/CE del Consiglio del 22 aprile 1999 concernente i valori limite di qualità dell'aria ambiente per il biossido di zolfo, il biossido di azoto, gli ossidi di azoto, le particelle e il piombo e della direttiva 2000/69/CE relativa ai valori limite di qualità dell'aria ambiente per il benzene ed il monossido di carbonio, Gazzetta Ufficiale n 87 del 13 Aprile 2002, Roma.

Piedmont Regional Law 43/2000. (2000). *Disposizioni per la tutela dell'ambiente in materia di inquinamento atmosferico. Prima attuazione del Piano regionale per il risanamento e la tutela della qualità dell'aria*. Piedmont Regional Council, Turin, Italy. Retrieved September 2018, from www.regione.piemonte.it.

Rapport, D. J., Howard, J., Lannigan, R., & Mccauley, W. (2003). Linking health and ecology in the medical curriculum. *Environment International, 29*(2–3), 353–358.

Ritz, B., Yu, F., Fruin, S., Chapa, G., Shaw, G. M., & Harris, J. A. (2002). Ambient air pollution and risk of birth defects in Southern California. *American Journal of Epidemiology, 155*(1), 17–25.

Roy, B. (1993). Decision science or decision-aid science. *European Journal of Operational Research, 66*, 184–204.

Roy, B. (1996). *Multicriteria methodology for decision aiding*. Dordrecht: Kluwer Academic Publishers.

Sunyer, J., Basagana, X., Belmonte, J., & Anto, J. M. (2002). Effect of nitrogen dioxide and ozone on the risk of dying in patients with severe asthma. *Thorax – International Journal of Respiratory Medicine, 57*(8), 687–693.

Tankersley, C. G., Georgakopoulos, D., Tang, W. Y., Abston, E., Bierman, A., & Sborz, N. (2012). Effects of ozone and particulate matter on cardiac mechanics: Role of the atrial natriuretic peptide gene. *Toxicological Sciences, 131*(1), 95–107.

U.S. EPA (Environmental Protection Agency). (1999). *Guideline for reporting of daily air quality – Air Quality Index*. EPA-454/R-99-010, North Carolina.

U.S. EPA (Environmental Protection Agency). (2006). *Guideline for reporting of daily air quality – Air Quality Index*. EPA-454/B-06-001, North Carolina.

U.S. EPA (Environmental Protection Agency). (2016). *Technical assistance document for the reporting of daily air-quality index (AQI)*. EPA-454/B-16-002, North Carolina.

UNDP (United Nations Development Programme). (1990). *Human development report 1990*. Oxford: Oxford University Press.

UNDP (United Nations Development Programme). (1997). *Human Development Report 1997*. Oxford: Oxford University Press.

UNDP (United Nations Development Programme). (2003). *Human Development Report 2003*. Oxford: Oxford University Press.

UNDP (United Nations Development Programme). (2018). *Human Development Report 2016*. Report Technical Notes, New York. Retrieved September 2018, from http://hdr.undp.org/en/2016-report.

Vansnick, J. C. (1986). De Borda et Condorcet à l'aggregation Multicritere. *Ricerca Operativa, 40*, 7–44.

Vincke, P. (1992). *Multi-criteria decision aid*. New York: Wiley.

Yang, I. A., Holz, O., Jörres, R. A., Magnussen, H., Barton, S. J., Rodríguez, S., Cakebread, J. A., Holloway, J. W., & Holgate, S. T. (2005). Association of tumor necrosis factor – Polymorphisms and ozone-induced change in lung function. *American Journal of Respiratory and Critical Care Medicine, 171*, 171–176.

Zarnowsky, F. (1989). *The decathlon – A colourful history of track and field's most challenging event*. Champaign: Leisure Press.

3 From Measurement Theory to Indicator Theory

Abstract

Although the focus of the monograph is on indicators, this chapter "takes a step back", (re)analysing the concept of *measurement* in the broadest sense. Precisely, the first part of the chapter recalls the historical evolution of this concept, focusing on *physical* measurements and their unification. Subsequently, the attention shifts on the *theory of measurement* by Stevens and several other measurement theorists, who had the great merit of extending the concept of measurement beyond physical quantities, identifying important properties, implications, and (mis)uses. Particular attention is paid to the concept of *meaningfulness* of statements using measurement scales.

Then, the rest of the chapter develops an original *theory of indicator*, showing that measurements can be seen as "special" indicators. The concept of *non-uniqueness* of representation by means of indicators is also explained. The description is supported by several practical examples.

3.1 Historical Evolution of the Concept of Measurement

Definition 3.1 Physicists traditionally consider the *measurement* as "a process by which one can convert physical parameters to meaningful numbers" (ISO/IEC GUIDE 99:2007 2007; JCGM 100:2008 2008).

Definition 3.2 A *measuring instrument* is a "device for determining the value or magnitude of a quantity/variable" (JCGM 100:2008 2008).

Definition 3.3 A *unit of measurement* is the "standard measure of the physical quantity of interest", i.e., the number of times the unit occurs in any given amount of the same quantity is the measurement itself (JCGM 100:2008 2008). Without the unit, the measurement number has no physical meaning.

F. Franceschini et al., *Designing Performance Measurement Systems*, Management for Professionals, https://doi.org/10.1007/978-3-030-01192-5_3

Until the eighteenth century there was no unified measurement system. On the contrary, thousands of alternative units of measurement were used around the world. Many were simply borrowed from human morphology: e.g., the digit, the hand, the foot, the cubit, the pace, the fathom, etc. These units usually varied from one town to another, from one occupation to another, as well on the type of objects to be measured. Apart from representing a source of error and fraud in commercial transactions, this situation also put a check on the development of science. With the expansion of industry and trade, there was an increasing need for unification.

France was the first nation to do anything concrete to remedy this situation. Precisely, politicians and scientists did their best to produce an invariable measure by comparison with a standard borrowed from a natural phenomenon, i.e., the universal standard that Condorcet had dreamed of as far back as 1775, which would not be based on any national vanity and which could be used by all foreign nations.

The climate of reform which followed the French revolution precipitated the choice of a standard. On 16 February 1791, following a proposal by Jean-Charles de Borda, a commission was set up to bring in a uniform system of measurement. On 26 March 1791, this commission defined the *metre* as being equal to the ten millionth part of one quarter of the terrestrial meridian. The metre materialised the idea of a "unit which in its determination was neither arbitrary nor related to any particular nation on the globe". Once the basic measurement unit had been determined, all that had to be done now was "just" establishing all the other resulting measurement units: the square metre and the cubic metre, the litre, the gram, etc.

The decimal metric system was introduced on 7 April 1795 by the French law "on weights and measures". This caused a major upheaval in everyday life. Decimalisation also brought a real revolution in the calculation of areas and volumes. Conversion from a multiple to a sub-multiple unit in area, and *vice versa*, simply consists of moving the decimal point two places, or three places for volume.

To determine the unit of mass, the commission preferred water to any other body such as mercury or gold, due to the ease of obtaining water and distilling it. The kilogram was defined as being equal to the mass of a cubic decimetre of water at a given temperature.

Both simple and universal, the decimal metric system started to spread outside France. The development of railways, the growth of industry and the increasing number of exchanges all required accurate units of measurement. More and more countries gradually adopted the metric system. Nevertheless, these countries were dependent on France whenever exact copies of the metre and kilogram standards were required. This subordination to France, together with the lack of uniformity in making copies, was likely to jeopardise the desired unification.

To overcome these difficulties, the *Bureau International des Poids et Mesures* (BIPM) was founded in 1875, during the diplomatic conference of the metre, which led, on 20 May 1875 to the signature of the treaty known as the Metre Convention by the plenipotentiaries of 17 States. The BIPM's initial mission was to set up the metric system throughout the world by constructing and maintaining new prototypes of the metre and the kilogram, comparing the national standards with these prototypes and perfecting the measurement methods in order to promote metrology in all fields. The BIPM progressively focused on the study of metrological problems and physical constants.

The *International System of Units* (SI), successor of the metric system, was officially founded in 1960 following a resolution made in the 11th *Conférence Générale des Poids et Mesures* (CGPM). All units of measurement can be reduced to a small number of fundamental standards with this system, which dedicates the necessary care to continuously improve their definition. There are two types of SI units, *basic* units and *derived* units. The seven SI basic units are *length*, *mass*, *time*, *electric current*, *thermodynamic temperature*, *amount of substance* and *luminous intensity*. Derived units are obtained by combination of basic units, through mathematical operations of multiplication and division (see the scheme in Fig. 3.1).

The definitions of the SI basic units, the measurement methods and the standards themselves undergo constant progress and are permanently renewed. The work on fundamental standards, carried out in particular by national metrology laboratories, will probably never end.

3.2 Stevens' Theory of Scales of Measurement

Although the SI played and will continue to play a fundamental role in the (re) definition of methods to measure physical quantities, it does not contemplate other (non-physical) quantities. Similarly, the Definition 3.1 (of measurement) is not exhaustive, since it is limited to measurements of physical quantities only. Yet the real world includes a great amount of measurements that are not covered by SI: e.g., the risk of a financial transaction, the spreadability of butter, the driveability of a car, the aggressiveness of a degenerative disease, the IQ of an individual, etc.

A more general measurement theory needs to be developed (Scott and Suppes 1958; Torgerson 1958; Churchman and Ratoosh 1960; Finkelstein 1982; Roberts 1979). One of the most important pioneers of measurement theorists is probably the Harvard psychologist Stanley Smith Stevens, who gave the following definition of measurement: "A *measurement*, in the broadest sense, is defined as the assignment of numerals to objects or events according to rules [. . .] The fact that numerals can be assigned under different rules leads to different kinds of scales and different kinds of measurement" (Stevens 1946). This definition will be later revised by the theorists of the so-called *representation theory of measurement*, discussed in Sect. 3.3.

Stevens coined the terms *nominal*, *ordinal*, *interval*, and *ratio* to describe a hierarchy of possible measurement scales and classified statistical procedures according to the scales for which they were "permissible." This taxonomy was subsequently adopted by several important statistics books and has thus influenced the statistical reasoning of a generation. Although Stevens's categories have been criticized by some statisticians because of their (alleged) incompleteness, they still are a consolidated reference for many books.

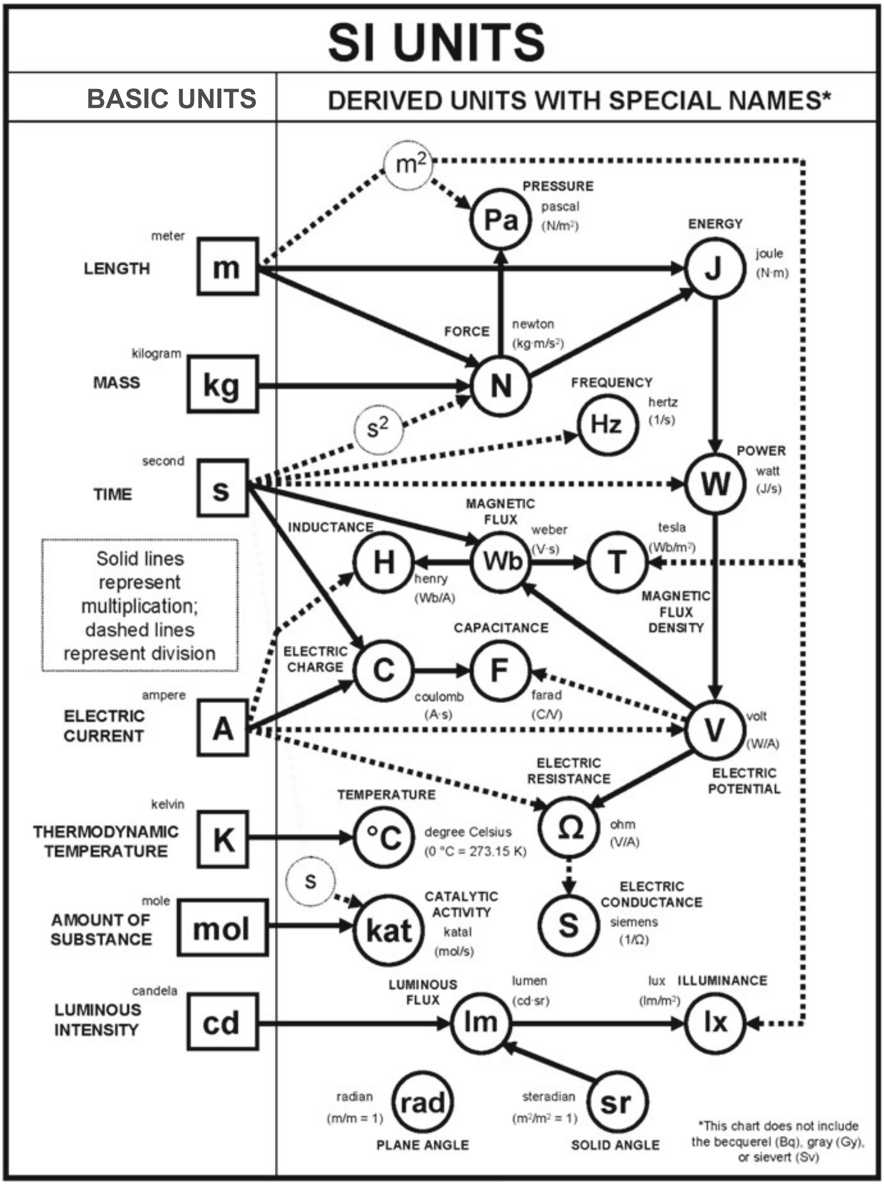

Fig. 3.1 Scheme of basic and derived units, according to the SI (adapted from U.S. Metric Association, www.us-metric.org [retrieved on September 2018]). With permission

In his seminal paper "On the theory of scales of measurement" Stevens (1946, 1959) carefully describes the four types of scales (see Table 3.1). A summary of this description is contained in the next four subsections.

Table 3.1 Classification of measurements depending on their scale types

Scale type	Basic relations among objects	Admissible scale transformations	Permissible statistics (invariantive)	Examples
Nominal	Equality ($=$, $\neq$)	Permutation function (one-to-one substitution)	Number of cases, mode, contingency correlation	Eye colour, place of birth, etc.
Ordinal	Order ($>$, $<$)	Monotonic increasing function	Median, percentiles	*Mohs* surface hardness, military rank, etc.
Interval	Equality of intervals or differences ($+$, $-$)	Linear increasing function: $\Phi(x) = a{\cdot}x + b$, being $a > 0$	Mean, standard deviation, rank-order correlation, product-moment correlation	Temperature in °C, calendar time, etc.
Ratio	Equality of ratios ($\times$, $/$)	Similarity: $\Phi(x) = a{\cdot}x$, being $a > 0$	All statistics permitted for interval scales plus the following: geometric mean, harmonic mean, coefficient of variation, logarithms	Temperature in K, mass, age, number of children, etc.
Absolute[a]	Idem as above	Identity: $\Phi(x) = x$	Idem as above	Any counting of items

Adapted from Stevens (1946, 1959). The columns on the *basic relations among objects* and the *permissible statistics* are cumulative, i.e., each item also includes those reported in the rows above it. Conversely, the column on *the admissible scale transformations* is inversely-cumulative, i.e., each item also includes those reported in the rows below it

[a]Special ratio scale that is used whenever counting items

3.2.1 Nominal Scale

The nominal scale differentiates between objects based only on their names or categories and other qualitative classifications they belong to. For example, consider a morphological characteristic of biological species (such as skin and eye colour), the sexual orientation of individuals (e.g., *heterosexuality*, *homosexuality*, *bisexuality*, *asexuality*, etc.), or the nationality of individuals.

The numbers in nominal measurement are assigned as labels and have no specific numerical value or meaning. No form of arithmetic computation (e.g., +, −, ×, / etc.) may be performed on nominal measurements. For this reason, the nominal scale is considered as the least powerful measurement scale, from a statistical point of view.

Mathematical operations Equality and other operations that can be defined in terms of equality, such as inequality and category membership, are the only non-trivial operations that generically apply to objects of the nominal type.

Admissible scale transformations Since the purpose of categorizing objects is just as well served when any two designating categories are interchanged, this scale form remains invariant under the general substitution or permutation function.

Measures of central location The mode, i.e. the most common category, is allowed as the measure of central location for the nominal type. On the other hand, the median, i.e. the middle-ranked item, makes no sense for the nominal type of data, since ranking is meaningless for the nominal scale.

3.2.2 Ordinal Scale

The ordinal scale allows for rank order (e.g., 1st, 2nd, 3rd, etc.) by which objects can be sorted, but still does not allow for relative degree of difference between them. Examples include data consisting of a spectrum of ordered values or categories, such as *general*, *brigadier*, *colonel*, *commandant*, *captain*, *officer* etc., when measuring military ranks. Other examples include Intelligence-Quotient (IQ) scores, which reflect an ordinal scale in which all scores are meaningful for comparison only; there is no absolute zero and a ten-point difference may carry different meanings at different points of the scale.

Admissible scale transformations Any series of symbols/numbers can be assigned to objects/categories, as long as they preserve their order relations (e.g., "object/category x_1 is greater than object/category x_2"). For this reason, *strictly monotonic increasing* transformations are admissible.

Measures of central location The median (or middle-ranked) object's category is allowed as the measure of central location; however, the mean (or average) value is not allowed. The mode is allowed.

Measures of statistical dispersion Percentile distance, interquartile range, etc.

3.2.3 Interval Scale

The interval scale allows for the degree of difference (or "distance") between objects, but not the ratio between them. Examples include temperature with the Celsius scale, which has two defined points (the freezing and boiling point of water at specific conditions) and then separated into 100 intervals, date when measured from an arbitrary epoch (e.g., *Anno Domini*, *Rosh haShana*, *Anno Hegirae*, etc.), location in Cartesian coordinates, and direction measured in degrees from true or magnetic north.

Ratios are not meaningful since 20 °C cannot be said to be "twice as hot" as 10 °C, nor can multiplication/division be carried out between any two dates directly. However, ratios of differences can be expressed (e.g., one difference can be twice another).

Admissible scale transformations Any positive linear transformation of the form $\Phi(x) = a \cdot x + b$ (being $a > 0$) will preserve the interval relations of objects (e.g., "the difference between object x_1 and object x_2 is larger than twice the difference between object x_3 and object x_4").

Measures of central location and statistical dispersion The mode, median, and arithmetic mean are allowed to measure central location of interval variables, while measures of statistical dispersion include range and standard deviation. Since one can only divide by differences, one cannot define measures that require some ratios, such as the coefficient of variation.

3.2.4 Ratio Scale

The ratio scale takes its name from the fact that measurement is the estimation of the ratio between a magnitude of a continuous quantity and a unit magnitude of the same kind. A ratio scale possesses a meaningful (unique and *non-arbitrary*) zero point. Most measurement in the physical sciences and engineering is done on ratio scales; e.g., consider the SI units in Fig. 3.1. In contrast to interval scales, ratios are meaningful because having a non-arbitrary zero point makes it meaningful to say, for example, that one object has "twice the length" of one other. Very informally, ratio scales can be described as specifying "how much" of something (i.e. an amount or magnitude) or "how many" (a count). The Kelvin temperature scale is a ratio scale because it has a unique, non-arbitrary zero point called *absolute zero* (i.e., $-273.15\ °C$).

Admissible scale transformations Any "similarity" transformation of the form $\Phi(x) = a \cdot x$ (being $a > 0$) will preserve the ratio relations of objects (e.g., "the ratio between object x_1 and object x_2 is equal to a certain value").

Measures of central location and statistical dispersion The geometric mean and the harmonic mean are allowed to measure the central location, in addition to the mode, median, and arithmetic mean. The "studentized" range and the coefficient of variation are allowed to measure statistical dispersion. All statistical measures are allowed because all necessary mathematical operations are defined for the ratio scale.

Note on absolute scales The so-called *absolute scale* is a special ratio scale which is used whenever counting items. Apart from a non-arbitrary zero, this scale also includes a non-arbitrary unit, corresponding to the single item. The application of transformations $\Phi(x) = a \cdot x$ (being $a > 0$) to absolute scales—although being admissible for ratio scales in general—would "distort" the unit, leading to the loss of its physical meaning. In this case, the only admissible scale transformation would be the "identity" $\Phi(x) = x$ (Roberts 1979).

3.2.5 Comments on Stevens' Scale Types

Stevens' scale types are ordered from *less* powerful to *more* powerful. The "power" is understood from the point of view of the possible relations among objects: in fact, the more powerful scales admit more relations than the less powerful ones (see Table 3.1).

On the other hand, the admissible transformations of a more powerful scale represent a subset of the transformations of a less powerful scale; e.g., the transformations related to a ratio scale (i.e., $\Phi(x) = a{\cdot}x$, being $a > 0$) represent a special case of those related to an interval scale (i.e., $\Phi(x) = a{\cdot}x + b$, being $a > 0$).

In addition, the four Stevens' scale types can also be aggregated into two families:

- **Categorical scales**, including *nominal* and *ordinal* scales, as they include a finite number of categories (or levels), which are generally associated with corresponding numerical values. While categories are unordered for nominal scales, they are ordered for ordinal scales. However, for both these scales the notion of difference (or distance) between objects is meaningless.
- **Cardinal scales**, including: *interval* e *ratio* scales, i.e., two scale types for which the notion of difference (or distance) between objects is meaningful. While the zero point is arbitrary for interval scales, it is absolute for ratio scales.

Example 3.1 The Kano (1984) model is used in the quality engineering field to categorize the *attributes* (or *customer requirements*) of a product into five categories (see Table 3.2).

Let us now analyze the Kano model, in the light of the following questions:

(i) On which scale type (according to Stevens' theory) can the above categories be defined?

(ii) A company has to find the most suitable Kano category for the attribute "stable connectivity" of a smartphone. Through a questionnaire, 50 potential users are individually asked to categorize the attribute of interest (see results in Table 3.3). The resulting category (k) is determined by the (rounded) weighted average of the selected categories, according to the numerical conversion $B = 3$, $O = 4$, $E = 5$ (weights are given by the relevant frequencies); categories I and R are not considered because they are not selected by any user. We point out that this numerical conversion (see also Table 3.3) is purely conventional and therefore arbitrary.

What is the resulting category? Is the proposed aggregation method reasonable?

(iii) Referring to point (i), what are the appropriate central location indicators for this scale?

Table 3.2 Description of Kano categories (Kano 1984; Mikulić and Prebežac 2011)

Category	Description
(*B*) Basic (Must-be)	These are the requirements that the customers expect and are taken for granted. When done well, customers are just neutral, but when done poorly, customers are very dissatisfied. Kano originally called these "Must-be's" because they are the requirements that must be included and are the price of entry into a market.
(*O*) One-dimensional	These attributes are typically "spoken" and result in satisfaction when fulfilled and dissatisfaction when not fulfilled. In other words, they are standard characteristics that increase or decrease satisfaction by their degree (cost/price, ease of use, speed).
(*E*) Excitement (or attractive)	These attributes provide satisfaction when achieved fully, but do not cause dissatisfaction when not fulfilled. These are attributes that are not normally expected, for example, a thermometer on a package of milk showing the temperature of the milk. Since these types of attributes of quality unexpectedly delight customers, they are often unspoken.
(*I*) Indifferent	These attributes refer to aspects that are neither good nor bad, and they do not result in either customer satisfaction or customer dissatisfaction. For example, thickness of the wax coating on a milk carton; this might be key to the design and manufacturing of the carton, but consumers are not even aware of the distinction. It is interesting to identify these attributes in the product in order to suppress them and therefore diminish production costs.
(*R*) Reverse	These attributes refer to a high degree of achievement resulting in dissatisfaction and to the fact that not all customers are alike. For example, some customers prefer high-tech products, while others prefer the basic model of a product and will be dissatisfied if a product has too many extra features.

Table 3.3 Results of a questionnaire submitted to a sample of 50 potential users, to determine the Kano category of the attribute "stable connectivity" of a smartphone

Category (i)	Frequency (f_i)	Numerical conversion (n_i)
(*B*) Basic	22	3
(*O*) One-dimensional	6	4
(*E*) Excitement	22	5
(*I*) Indifferent	0	Not applicable
(*R*) Reverse	0	Not applicable
	Sum: 50	

With reference to question (i), Kano categories can be defined on a *nominal* or (at most) *ordinal* scale, upon the (questionable/subjective) assumption that the order relation $B < O < E$ holds (e.g., when considering the degree of customer satisfaction).

Regarding question (ii), the resulting category (i.e., the category that is supposed to be most representative of the attribute of interest) is determined as follow:

$$n_k = \left[\sum_{i \in \{B,O,E\}} \left(\frac{n_i \cdot f_i}{\sum_{i \in \{B,O,E\}} f_i} \right) \right] = \left[\frac{3 \cdot 22 + 4 \cdot 6 + 5 \cdot 22}{22 + 6 + 22} \right] = 4$$
$$\Rightarrow \quad k = O, \tag{3.1}$$

where n_k is the numerical conversion of the resulting Kano category (k) and "[]" denotes the *nearest integer* operator.

Paradoxically, O is the category selected by the lowest number of users (i.e., 6 out of 50). This paradox is due to the fact that the aggregation based on the weighted average is not compatible with the scale type of Kano categories (i.e., nominal or ordinal). In fact, the use of the (weighted) average is applicable to interval or ratio scales only (see Table 3.1). Furthermore, the numerical conversion of (supposed) ordered categories with consecutive integers. The resulting aggregation is therefore not reasonable.

With reference to question (iii), appropriate central location indicators are: the *mode*, in the hypothesis that the scale type is nominal. We remark that the use of a weighted average (see Eq. 3.1) is not consistent with the aforementioned scale types.

3.2.6 Meaningfulness of Statements Concerning Scales

In general, the correct or incorrect use of a measurement scale depends on the *statements* (or assertions) regarding the data/objects under that scale and the relevant properties.

At the heart of Stevens' approach is the notion that statements should be invariant under change of scale, that is, they should not be scale-dependent. This notion of *invariance* is the motivation for the concept of *meaningfulness*, a concept first made precise by several measurement theorists, such as Luce, Suppes, Zinnes, Pfanzagl, Roberts, etc. (Krantz et al. 1971, 1989, 1990).

Definition 3.4 A statement using scales is called *meaningful* "*if its truth or falsity is unchanged whenever any scale is replaced by another acceptable scale*" (for instance by changing units) (Roberts 1979); otherwise, it is called *meaningless*.

For example, consider the statement: "I am twice as tall as the Obelisk of Theodosius in Istanbul" (around 20 m high!); this statement is clearly false for all scales of height, i.e., it is false whether we use inches, feet, meters, or any other unit. Thus, the statement is meaningful because its falsity is (in this case) independent of the scales used. Note that meaningfulness is not the same as *truth*. It has to do with the "compatibility" of a statement with a particular measurement scale.

As seen, for each scale of measurement, Stevens identified a class of *admissible transformations*, i.e., a function Φ that, when applied to the scale of interest produces another acceptable scale. For instance, in the case of height, any acceptable scale can be transformed into another acceptable scale through the product by a positive constant: $\Phi(x) = a \cdot x$, being $a > 0$ (cf. Table 3.1). Thus, we transform from inches to meters by multiplying by 0.0254 and from inches to feet by multiplying by 1/12. Multiplication by a positive constant defines an admissible transformation. All admissible transformations can be described this way.

From the perspective of admissible scale transformations, the Definition 3.4 (of meaningfulness) can be modified as:

Definition 3.5 "*A statement using scales is called meaningful if its truth or falsity is unchanged when all scales in the statement are transformed by admissible transformations*" (Marcus-Roberts and Roberts 1987); otherwise, it is called *meaningless*.

One of the main obstacles to applying the Stevens' theory of scales of measurement is that we frequently do not know the class of admissible transformations. The *representational theory of measurement*, which will be recalled in Sect. 3.3, develops some theorems which lead to the specification of the scale type under certain axioms (Roberts 1979). However, in practice, the user of the theory of scale type tries to specify the admissible transformations on the basis of some procedure used in the measurement or on the basis of intuition.

One of the tasks of measurement theorists over the years has been to characterize what statements involving scales of measurement are meaningful. The following are some sample results:

- If some objects are defined on a *ratio* scale, then it is meaningful to say that one object is so-and-so many times as big as another (e.g., mass x_1 is k times as high as mass x_2).
- If some objects ($x_1, x_2, ...$) are defined on an *interval* scale, then it is meaningful to say that $(x_1 - x_2)$ is greater than $(x_3 - x_4)$ (e.g., the difference between the highest and lowest temperatures today is greater than the corresponding difference yesterday).
- If some objects ($x_1, x_2, ...$) are defined on an *ordinal* scale, then it is meaningful to say that $x_1 > x_2$ (e.g., one object rated 4 is better than another one rated 3 on a five-level ordinal scale).
- If x_1 and x_2 are two objects defined on a *nominal* scale, then it is meaningful to say that $x_1 = x_2$ (e.g., the eye colour of one individual is equal to that of one other).

In general, the meaningfulness of a statement can be formally demonstrated through an (analytical) proof that its truth/falsity is invariant under whatever admissible transformation. We point out that a specific example of the invariance of the statement under one of the possible admissible transformations does not represent a general proof of meaningfulness (*one swallow doesn't make summer*!).

On the other hand, the meaninglessness of a statement can be formally demonstrated through either a general (analytical) proof or a single counter-example, which shows the non-invariance of the statement under a specific admissible scale transformation (see also the scheme in Fig. 3.2).

Fig. 3.2 Scheme of the possible ways to formally demonstrate the meaningfulness/meaninglessness of a statement. The symbols "tick" and "cross" indicate respectively the validity/invalidity of a certain way of demonstration

	MEANINGFULNESS	MEANINGLESSNESS
GENERAL (ANALYTICAL) PROOF	✓	✓
SINGLE (COUNTER-)EXAMPLE	✗	✓

Having demonstrated that a statement is meaningful/meaningless under a certain scale type, it is possible to make deductions concerning other scale types, according to the following rules.

Rule 3.1 If a statement is meaningful under a certain scale, it will also be meaningful under a more powerful scale. In fact, if a statement is invariant under the admissible transformations of a certain scale type, it will also be invariant under the transformations of more powerful scale types, since the latter transformations represent a subset of the former ones (see Table 3.1).

Rule 3.2 If a statement is meaningless under a certain scale, it will also be meaningless under a less powerful scale. In fact, if a statement is non-invariant under the admissible transformations of a certain scale type, it will also be non-invariant under the transformations of less powerful scale types, since the latter transformations include the former ones (which represent a special case).

Example 3.2 Let us prove that the statement "$x_1 > x_2 + x_3 + x_4$" is meaningful for objects (i.e., x_1, x_2, x_3 and x_4) defined on a *ratio* scale. By applying a general admissible transformation for ratio scales (i.e., $\Phi(x) = a{\cdot}x$, being $a > 0$) to all four objects, we obtain:

$$\Phi(x_1) > \Phi(x_2) + \Phi(x_3) + \Phi(x_4), \tag{3.2}$$

from which:

$$a \cdot x_1 > a \cdot x_2 + a \cdot x_3 + a \cdot x_4 \quad \Rightarrow \quad x_1 > x_2 + x_3 + x_4. \tag{3.3}$$

We note that the statement with "transformed" objects degenerates into the initial one. This proves that the truth/falsity of the statement is invariant under whatever admissible scale transformation.

Table 3.4 (a) Annual production (in millions of cars) of five car manufacturers (A, B, C, D and E), in two consecutive years. (b) Data normalized through the *min-max normalization*

	(a) Initial data (10^6)					(b) Normalized data (%)				
Year	A	B	C	D	E	A	B	C	D	E
2015	3	5	9	6	2	14.3	42.9	100.0	57.1	0.0
2016	2	4	7	6	1	16.7	50.0	100.0	83.3	0.0

Example 3.3 Let us prove that the statement "$x_1 > x_2 + x_3 + x_4$" is meaningless for objects defined on an *interval* scale. By applying a general admissible transformation for interval scales (i.e., $\Phi(x) = a{\cdot}x + b$, being $a > 0$) to all four objects, we obtain:

$$(a \cdot x_1 + b) > (a \cdot x_2 + b) + (a \cdot x_3 + b) + (a \cdot x_4 + b) \\ \Rightarrow x_1 > x_2 + x_3 + x_4 + \left(\frac{2 \cdot b}{a}\right). \tag{3.4}$$

The statement with "transformed" objects is different from the initial one. In fact, depending on the specific x_1, x_2, x_3, x_4 values, and those of the parameters (i.e., a and b) of the scale transformation, the truth/falsity of the initial statement can be subverted. E.g., considering $x_1 = 7$, $x_2 = 1$, $x_3 = 2$, $x_4 = 3$, and $a = b = 1$, the statement is true before the transformation (i.e., 7>6) and false after the transformation (i.e., 7>8). This specific counter-example can be considered as a valid proof that the statement of interest is meaningless under interval scales and therefore also under ordinal and nominal scales (cf. Rule 3.2).

Example 3.4 Table 3.4a reports data concerning the annual production (in millions of cars) of five car manufacturers (A, B, C, D and E), in two consecutive years (2015 and 2016). To make comparison easier, these data are normalized through the so-called min-max normalization, i.e., $y = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \cdot 100$, obtaining the data reported in Table 3.4b. For example, for manufacturer A in the year 2015: $y = [(3-2)/(9-2)]{\cdot}100 = 14.3\%$.

We now critically analyse the proposed normalization, answering the following questions:

(i) Is this normalization consistent with the scale type of the initial data (in Table 3.4a)?

(ii) Considering the normalized data (in Table 3.4b), are the following statements meaningful?

- "In 2016, the performance of C is twice that of B";
- "In 2015, the performance gap between C and B is equal to that between D and E";
- "regarding B, the difference in performance between 2015 and 2016 is of +7.1 percent points".

It can be noticed that initial data (i.e., number of cars annually produced) are defined on a ratio scale (to be precise an *absolute* scale, cf. Sect. 3.2.4). The *min-max normalization* can be interpreted as a specific scale transformation:

$$y = \Phi(x) = a \cdot x + b \qquad (a > 0)$$
$$\Rightarrow \qquad y = \left(\frac{100}{x_{\max} - x_{\min}}\right) \cdot x - \left(\frac{x_{\min} \cdot 100}{x_{\max} - x_{\min}}\right), \tag{3.5}$$

i.e., an (increasing) linear function, which is therefore permissible for *interval/ ordinal/nominal* scales (see Table 3.1). The application of this transformation to objects defined on ratio scales may lead to "degrading" them to objects defined on interval scales, with consequent loss of meaning of their possible ratio relations. This concept will be clarified later on.

The first statement is true when considering the normalized data ($y_C = 100$ and $y_B = 50$) but it is false when considering the initial data ($x_C = 7$M and $x_B = 4$M). This apparent paradox is due to the scale degradation (from *ratio* to *interval*), due to the min-max normalization. This proves that this statement is meaningless.

The second statement concerns the so-called "distance" (or difference/interval) between objects. We observe that it is true when considering both the initial data ($x_C - x_D = x_B - x_E = 3$M) and the normalized ones ($y_C - y_D = y_B - y_E = 43$). The reason is that, although the initial ratio scale has been "degraded" to an interval scale, it still preserves the distances among objects (up to a positive scale factor). A general demonstration—applying a permissible transformation like $y = \Phi(x) = a{\cdot}x + b$, with $a > 0$—is left to the reader.

The third statement is true when considering normalized data (i.e., $y_{B,2016} - y_{B,2015} = 50 - 42.9 = 7.1$) but it becomes false when considering initial data (i.e., $x_{B,2016} - x_{B,2015} = 4\text{M} - 5\text{M} = -1\text{M}$). This (apparent) paradox proves that this statement is meaningless. Using a metaphor, comparing normalized data related to different years is like comparing "apples with oranges"; the representation in Fig. 3.3 clarifies this concept: the scales related to the normalized data are characterized by an arbitrary zero, which depends on the position of x_{min} in the reference year, and an arbitrary unit, which depends on the difference ($x_{max} - x_{min}$) in the reference year.

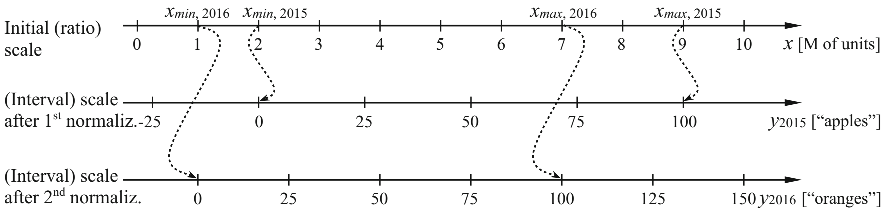

Fig. 3.3 Scheme of the measurement scales related to the data in Table 3.4, before and after the (min-max) normalization. The figure highlights the "misalignment" between the scales related to 2 years of interest (i.e., y_{2015} and y_{2016})

3.2.7 Statements Using (Non-)permissible Statistics

It can be demonstrated that statements involving non-permissible statistics (i.e., measures of central-tendency and statistical-dispersion) are meaningless (Marcus-Roberts and Roberts 1987). Below is an example that clarifies this concept.

Example 3.5 Suppose we measure the average (arithmetic mean height) of all students in each of two schools. Is it meaningful to assert that the average height of students in the first school is larger than the average height of students in the second school? If we let A_i be the height of a generic i-th student in school A and B_i be the height of a generic i-th student in school B, then we are making the statement

$$\frac{1}{n} \cdot \sum_{i=1}^{n} A_i > \frac{1}{m} \cdot \sum_{i=1}^{m} B_i, \tag{3.6}$$

where $A_1, A_2, \ldots, A_n$ are the (n) students in school A and $B_1, B_2, \ldots, B_m$ the (m) students in school B. Now height is measured on a *ratio* scale and admissible transformations involve multiplication by a positive constant a (i.e., $\Phi(x) = a{\cdot}x$, being $a > 0$):

$$\frac{1}{n} \cdot \sum_{i=1}^{n} a \cdot A_i > \frac{1}{m} \cdot \sum_{i=1}^{m} a \cdot B_i. \tag{3.7}$$

Since a is positive, Eq. (3.7) can be reduced to Eq. (3.6) (multiplying both terms by the positive quantity $1/a$), proving the meaningfulness of the statement of interest. The conclusion of meaningfulness still holds if we compare averages under *interval* scales (the proof is left to the reader).

However, meaningfulness can fail if we compare averages under *ordinal* scales. To demonstrate the latter assertion, suppose we measure each student's height on a five-level scale (l = *poor*, 2 = *fair*, 3 = *good*, 4 = *very good*, 5 = *excellent*). In such a scale, numbers do not mean anything *per se*, only their order matters. If school A has three students and their scores are 1, 3, and 5, and school B has three students and their scores are 2, 4, and 4, then school A has an average score of 3 and school B has an average score of 3.33, which is larger (statement is false). Now the five-level scale is just ordinal, and we could just as easily use the scale obtained from it by applying any strictly monotonic increasing transformation to the numbers used, for instance replacing l by 30, 2 by 40, 3 by 65, 4 by 75, and 5 by 100 (see the graphical representation of this transformation in Fig. 3.4). The new scale conveys the same information: it tells the order of the five categories. But on the new scale, school A has scores of 30, 65, and 100, which average out to 65, while school B has scores of 40, 75, and 75, which average out to 63.33. Now school B has a lower average (statement is true). Thus, comparison of arithmetic means with ordinal scales can be meaningless. In sum, we can compute means with ordinal data, but we can get misled by comparing them.

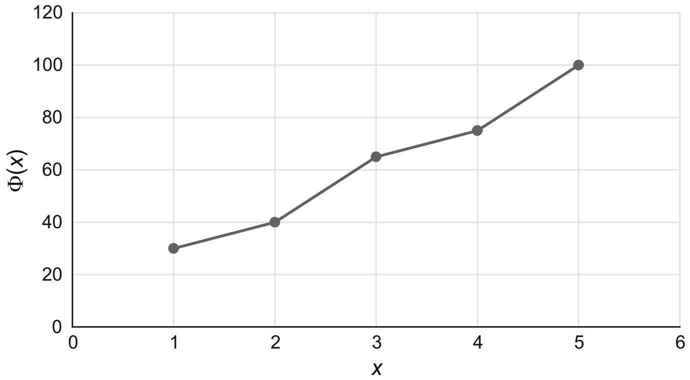

Fig. 3.4 Example of strictly monotonic increasing transformation $\Phi(x)$

Regarding central-tendency measures, it can also be shown that—under a scale for which they are permissible—the following condition holds:

$$\Phi[f(x_1, x_2, \ldots, x_n)] = f[\Phi(x_1), \Phi(x_2), \ldots, \Phi(x_n)], \tag{3.8}$$

being f the central-tendency measure and Φ the (permissible) scale transformation. We specify that the condition in Eq. (3.8) does not apply to (permissible) measures of statistical dispersion.

Example 3.6 We now verify the following statements, checking the condition in Eq. (3.8).

(i) Assuming that M_1 and M_2 are two objects defined on an *interval* scale, their arithmetic mean is a permissible central-tendency measure.
(ii) Assuming that M_1 and M_2 are two objects defined on an *ordinal* scale, their arithmetic mean is a non-permissible central-tendency measure.

The arithmetic mean (f) of the two objects is expressed as:

$$\bar{M} = f(M_1, M_2) = \frac{M_1 + M_2}{2}. \tag{3.9}$$

Let us assume that M_1 and M_2 are defined on an *interval* scale, which therefore admits the following (linear) transformation:

$$\Phi(x) = a \cdot x + b, \quad a > 0. \tag{3.10}$$

Applying this transformation to the right-hand member of Eq. (3.9), we obtain:

$$f[\Phi(M_1), \Phi(M_2)] = \frac{(a \cdot M_1 + b) + (a \cdot M_2 + b)}{2} = \frac{a \cdot (M_1 + M_2)}{2} + \frac{2 \cdot b}{2} = a \cdot \bar{M} + b, \quad (3.11)$$

and applying it to the left-hand member, we obtain:

$$\Phi[f(M_1, M_2)] = \Phi(\bar{M}) = a \cdot \bar{M} + b. \quad (3.12)$$

This therefore proves that the condition in Eq. (3.8) holds (independently from the values of a and b) and therefore verifies the statement (i):

$$\Phi[f(M_1, M_2)] = f[\Phi(M_1), \Phi(M_2)]. \quad (3.13)$$

Let us now assume that M_1 and M_2 are defined on an *ordinal* scale, which therefore admits monotonic increasing transformations. For example, we consider the transformation $\Phi(x) = x^3$. Applying this transformation to the right-hand member of Eq. (3.8), we obtain:

$$f[\Phi(M_1), \Phi(M_2)] = \frac{M_1^3 + M_2^3}{2}, \quad (3.14)$$

and applying it to the left-hand member, we obtain:

$$\Phi[f(M_1, M_2)] = \Phi(\bar{M}) = \bar{M}^3 = \left(\frac{M_1 + M_2}{2}\right)^3. \quad (3.15)$$

In this case, the condition in Eq. (3.8) is not necessarily verified. For example, assuming that $M_1 = 3$ and $M_2 = 5$, we obtain:

$$\Phi[f(M_1, M_2)] = \Phi(4) = \left(\frac{3+5}{2}\right)^3 = 4^3 = 64 \neq f[\Phi(M_1), \Phi(M_2)] = \frac{\Phi(3) + \Phi(5)}{2} = \frac{3^3 + 5^3}{2} = 76. \quad (3.16)$$

This counter-example verifies the statement (ii): i.e., the arithmetic mean is not a permissible statistic for objects under an ordinal scale (and also under nominal scales, cf. Rule 3.2).

3.3 The "Representation Theory of Measurement"

The foundational work of Stevens (1946) provided a relevant classification of scales of measurement and rules for the use of statistical procedures, which established that certain statistics and certain statements were inappropriate for certain scales of measurement.

Different measurement theorists—e.g., Churchman, Luce, Pfanzagl, Ratoosh, Roberts, Suppes, Zinnes, etc.—have been stimulated by Stevens' theory to construct a more modern, general and in-depth theory, known as *representation theory of measurement* (Krantz et al. 1971, 1989, 1990; Churchman and Ratoosh 1960). This theory includes the following definition of measurement:

Definition 3.6 "Measurement is the assignment of numbers to properties of objects or events in the real world by means of an *objective* and *empirical* operation, in such a way as to describe them. The modern form of measurement theory is *representational*: numbers assigned to objects/events must represent the perceived relations between the properties of those objects/events" (Finkelstein and Leaning 1984).

It can be noted that a measurement is a representation of reality that requires two fundamental attributes: empiricity and objectivity (Finkelstein 1982; Cecconi et al. 2007; Franceschini et al. 2007).

Empiricity arises when a representation is the result of observations and not, for example, of a thought experiment. Furthermore, the concept of *property* measured must be based on empirically determinable relations and not, say, on convention" (Finkelstein 2003).

There is empiricity when the relations among the observed objects are "real" and can be observed without ambiguity. Empiricity also means that there is "*an objective rule for classifying some aspect of observable objects as manifestations of the property*" (Finkelstein 1982).

Objectivity concerns the results that the representation produces, "*within the limits of error independent of the observer*" (Finkelstein 1982). "*Experiments may be repeated by different observers and each will get the same result*" (Sydenham et al. 1989).

Paraphrasing the above considerations, measurement is an operation of objective description of reality: different measurements of the same entity (in the same operating conditions) should result in the same output, independently from subjects. We suppose there is no "error" and uncertainty in an ideal measurement process (i.e., the effect of environment and other influential variables is negligible). It is also an empirical operation: "*Measurement has something to do with assigning numbers that correspond to or represent or "preserve" certain observed relations*" (Roberts 1979).

Going into the *representation theory of measurement*, a measurement can be seen as a representation of an observable property/feature of some objects into a set of symbols/numbers (Roberts 1979; Finkelstein 2003). This representation should include four parts:

(1) **An empirical relational system**. Considering a certain property or attribute (for example the length) of a generic *i*-th object of interest, let a_i represent the individual *manifestation* of this property. Extending the concept to all the

examined objects, we define the set of all the possible manifestations of the property of interest as:

$$\mathrm{A} = \{a_1, \ldots, a_i, \ldots\} \tag{3.17}$$

and a family of empirical *relations* among the elements of A

$$\mathrm{R} = \{R_1, \ldots, R_m\}. \tag{3.18}$$

We point out that the possible relations are those already contemplated by Stevens: e.g., "a_1 is equal to a_2", "a_1 is greater than a_2", etc. Then **A** is a corresponding *empirical relational system*:

$$\mathbf{A} = \langle \mathrm{A}, \mathrm{R} \rangle. \tag{3.19}$$

(2) **A symbolic relational system**. Let Z represent a class of symbols (typically, but not necessarily, numbers):

$$\mathrm{Z} = \{z_1, \ldots, z_i, \ldots\}, \tag{3.20}$$

and P is a family of *relations* among the symbols of Z:

$$\mathrm{P} = \{P_1, \ldots, P_m\}, \tag{3.21}$$

then **Z** is a corresponding *symbolic relational system*:

$$\mathbf{Z} = \langle \mathrm{Z}, \mathrm{P} \rangle. \tag{3.22}$$

(3) **A representation condition**. Measurement is defined as an objective empirical operation such that the empirical relational system $\mathbf{A} = \langle \mathrm{A}, \mathrm{R} \rangle$ is mapped into the symbolic relational system $\mathbf{Z} = \langle \mathrm{Z}, \mathrm{P} \rangle$ (see Fig. 3.5) (Finkelstein 2003).

Two mapping functions are defined:

$$\mathrm{M} : \mathrm{A} \to \mathrm{Z} \quad \text{(homomorphism)} \tag{3.23}$$

and

$$\mathrm{F} : \mathrm{R} \to \mathrm{P} \quad \text{(isomorphism)}, \tag{3.24}$$

so that $\mathrm{M}(a_i) = z_i$ is the image in Z of a generic element a_i of A, and $\mathrm{F}(R_j) = P_j$ is the image, in P, of a generic *relation* R_j in R.

Focusing on M, this mapping is not a one-to-one function: separate but indistinguishable manifestations can be mapped into the same symbol; some practical situations in which this occurs are exemplified below.

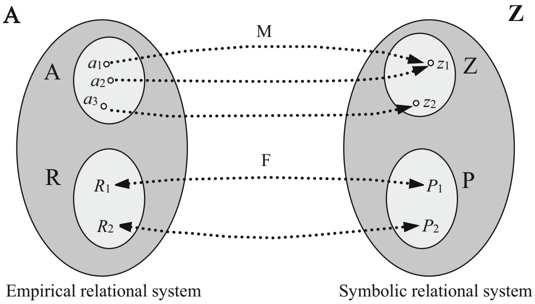

Fig. 3.5 Schematic representation of the concept of *measurement*. A measurement is a mapping of an *empirical relational system* (**A**) into the *symbolic relational system* (typically a *numerical* one) (**Z**) (Roberts 1979; Finkelstein 2003). A and R are respectively the set of the *manifestations* and the corresponding *relations* in **A**; Z and P are respectively the set of the *manifestations* and the corresponding *relations* in **Z**; M and F are the mapping functions from **A** to **Z**

Example 3.7 Let us consider the classification of living organisms into biological species: e.g., *bacteria*, *protozoa*, *chromista*, *plantae*, *fungi*, *animalia*. Obviously, some of the organisms examined (i.e., the empirical manifestations a_1, a_2, ...) may be classified into the same (nominal) categories (i.e., the symbols z_1, z_2, ...). For example, *escherichia coli* and *salmonella* are two different living organisms that are classified in the same category (i.e., *bacteria*).

Example 3.8 When measuring the mass of some objects, it could happen that different objects (i.e., empirical manifestations) may have the same identical mass (i.e., symbol).

Focusing on F, this other mapping is a one-to-one function: the relations among symbols should match the relations among empirical manifestations. Since the most commonly used symbols are numbers, typical relations are the mathematical ones.

The matching between empirical and symbolic relations (i.e., $F : R \rightarrow P$) entails that: (i) we should not introduce new relations among symbols, which do not reflect any existing relation among empirical manifestations (i.e., “promotion” of relations) or (ii) we should not omit to reflect some relations among empirical manifestations, through corresponding relations among symbols (i.e., loss/degrade of relations). According to Roberts (1979), “in measurement we start with an observed or empirical relational system and we seek a mapping to a numerical relational system which preserves all the relations and operations observed in the empirical one”. Similarly, Dawes and Smith (1985) state that “whatever inferences can be made in the symbolic/numerical relational system apply to the empirical one”.

Example 3.9 In the hypothesis that one wants to measure the mass of some objects, (i.e., a quantity defined on a *ratio* scale), the empirical relations of *equality*, *order*, *equality of intervals* and *equality of ratio*s must be reflected by correponding mathematical relations among the numbers (symbols) used to represent them.

The mapping functions (M and F) must be *objective*, i.e., they should lead to the same result, independently of the individual who performs the mapping.

(4) **Uniqueness condition**. The representation condition may be valid for more than one mapping function M. For a generic measurement, there is a *class of admissible transformations* from one scale to another scale for which the representation condition is valid (Finkelstein and Leaning 1984; Franceschini 2001; Finkelstein 2003). In other words, the representation provided by a measurement is not unique but there are many possible valid representations, which are connected to each other through the so-called *admissible scale transformations*, theorized by Stevens (1946) (cf. Table 3.1). For example, when measurements of height can be expressed in feet, inches, centimetres, or—more in general—acceptable (ratio) scales that are linked to each other through admissible scale transformations in the form $\Phi(x) = a{\cdot}x$ (being $a > 0$).

3.4 The Formal Concept of Indicator

This section, switches from the concept of measurement to that of indicator, providing a theoretical framework that will lead to the formulation of a (new) *representation theory of indicator*.

3.4.1 General Definitions

The concept of *indicator* is strictly related to the notion of *representation target*.

Definition 3.7 A *representation target* is a specific aspect of a *context* that we want to represent, in order to make evaluations, comparisons, predictions, decisions, etc.

E.g., for a supply chain (context), we can consider the service level or the customer satisfaction (representation targets).

For each representation target (Q) one can define a set of *indicators* (S_Q) capable of operationalizing it. Reversing the concept, we can formulate the following definition of indicator.

Definition 3.8 Indicators are conceptual tools that can *operationalize* a representation target, i.e., implementing a consistent representation of it.

We can define a generic set of n indicators as:

$$S_Q = \{I_1, I_2, \ldots, I_n\}_Q \qquad n \in \mathrm{N}, \tag{3.25}$$

being N the set of natural numbers {0, 1, 2, ...}. For example, if the *context* is the "logistic process" of a company and the representation target is the "classification of suppliers", two possible (pertinent) indicators are (1) "delivery time" and (2) "portion of defective products".

For a given representation target, a set of indicators cannot (unfortunately) be generated algorithmically (Roy and Bouyssou 1993). The selection of indicators is generally driven by the experience of indicator designers and their knowledge of the context of interest.

3.4.2 The "Representation Theory of Indicator"

Following a strict parallelism with the *representation theory of measurement*, an *indicator* can be considered as a mapping from an *empirical system* (the "real world") into a *symbolic system* (e.g., a *numeric system*). However, this mapping is more "relaxed" than for measurements, for two reasons:

1. The homomorphism between empirical manifestations and symbols (i.e., M: A→Z) does not necessarily have to be *objective*.
2. The isomorphism between empirical relations and symbolic relations (i.e., F: R→P) is not necessarily required.

In a nutshell, indicators perform a not-necessarily-objective homomorphical mapping of the elements of A into a set of symbols (contained in Z). Secondly, the mapping between empirical and symbolic relations is not relevant, e.g., (i) symbols can include "new" relations, which do not reflect any existing relation among empirical manifestations (i.e., "promotion" of relations) or (ii) some relations among empirical manifestations cannot be reflected by any specific relation among symbols (i.e., loss/degrade of relations).

Of course, the missing matching between empirical relations and symbolic relations is not desirable, as it can distort the representation. This point will be explored in the following chapter.

Example 3.10 The degree of satisfaction of a customer for a service—expressed on a five-level ordinal scale: 1-very low, 2-low, 3-intermediate, 4-high, 5-very high satisfaction—is an example of subjective mapping of an empirical manifestation. In fact, different custromers can map this degree of satisfaction in different scale categories.

Example 3.11 In Formula 1 races, the order of arrival of drivers (which obviously implies the existence of order relations between them) determines a consequent score (e.g., 25 points scored by the first finisher, 18 by the second, 15 by the third, and so on). Since scores are accumulated by drivers race after race, they are actually treated as *interval*-scale quantities (see Table 3.1). In other terms, the indicator "score of a driver" determines a promotion of the relations among symbols (i.e., scores) with respect to the relations among empirical manifestations (i.e., order of arrival of drivers).

Example 3.12 The qualifying session of Formula 1 races determines the drivers' starting position (ordinal scale) on the basis of the best lap-time (ratio scale). The ratio and interval relations of the best lap-times of drivers (i.e., empirical manifestations) are therefore not taken into account when determining the relevant starting positions (symbols). Consequently, the indicator "starting position of a driver" determines a degrade of the symbolic relations with respect to the empirical relations.

According to the proposed representational approach, indicators can be interpreted as representation models of (part of) the real world, through appropriate symbols/numbers, consistently with a certain representation target.

In addition, measurements may be interpreted as a subset of indicators. A basic difference between *measurements* and *indicators* is the way the relations of the *empirical systems* are mapped. *Indicators* do not require any isomorphism between empirical and symbolic *relations* (Eq. 3.22): i.e., the relations among symbolic manifestations (numbers) do not necessarily have to reflect the relations among empirical manifestations.

Another difference is that the mapping is *not necessarily objective* (i.e., it can change from subject to subject), generating subjective indicators. This means that, while a *measurement* is certainly an *indicator*, the opposite is not true (see Fig. 3.6).

Example 3.13 Let us consider the problem of choosing of a car. The customer preference is an indicator that maps an empirical system, which consists of different car models, into a symbolic system, which consists of corresponding preference ranks (1st, 2nd, 3rd, etc.). Since the ordinal relation among preference ranks does not reflect any objective (ordinal) relation among the real car models, this indicator is not a *measurement*. We remark that objectivity is a *sine-qua-non* condition for a measurement.

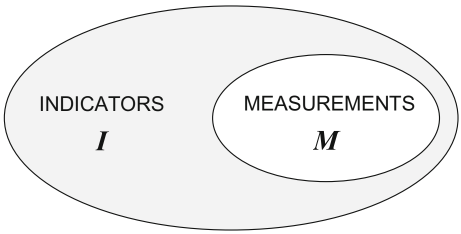

Fig. 3.6 Interpretation of *measurements* as a subset of *indicators*

3.5 Classification of Indicators

The following sub-sections provide a preliminary classification of indicators, dividing them into *objective/subjective* and *basic/derived*. This classification will be further developed in Chap. 4.

3.5.1 Objective and Subjective Indicators

Definition 3.9 *Objective* indicators objectively map empirical manifestations into symbolic ones, i.e., the mapping does not change from subject to subject. On the other hand, the mapping related to *subjective* indicators can change from subject to subject.

Since subjective indicators provide essential information on the behaviour and perceptions of individuals, they are often used in many disciplines in the area of *Social*, *Behavioural* and *Cognitive Sciences* (Nunally and Bernstein 1994; Arrow 1963). Most subjective indicators are expressed in the form of ordered scale categories (e.g., a five-level ordinal scale for evaluating the degree of customer satisfaction from a product: 1-very dissatisfied, 2-dissatisfied, 3-neutral, 4-satisfied, 5-very satisfied).

The conversion of scale categories into numbers is a common way to make the information practical to users. However – when the relations among symbolic manifestations do not match the relations among the empirical manifestations – the conversion may distort the analysis (Roberts 1979; Franceschini and Romano 1999; Narayana 1977; Franceschini 2001).

Let us make a brief digression on subjective indicators, distinguishing them according to their (non) *empiricity*. The *preference* is neither empirical nor objective. Preferences are, by definition, subjective and conflicting. For example, the voting preference does not necessarily have any empirical reference but may depend on unobservable factors.

The *evaluation* is somewhere between measurement and preference. It is not objective because evaluations are individual perceptions, performed without the use of a univocal instrument of measurement. Nevertheless, the evaluation is an operation that *wants* to be empirical as it should be derived from something that can be observed in a real experiment. Table 3.5 distinguishes between the three types of indicators: measurements, evaluations and preferences. For further details on the above distinction, see Cecconi et al. (2006, 2007), Franceschini et al. (2007).

Table 3.5 Subdivision of indicators into measurements, evaluations and preferences (Cecconi et al. 2006, 2007; Franceschini et al. 2007)

	Objective	Empirical
Measurement	Yes	Yes
Evaluation	No	Yes
Preference	No	No

3.5.2 Basic and Derived Indicators

Definition 3.10 An indicator is *basic*, if it is obtained through a direct observation of an empirical system. Examples of basic indicators are the number of defectives in a production line, the number of manufactured parts, the time period between two events, etc.

Definition 3.11 An indicator is *derived* if it is obtained through the manipulation/synthesis/aggregation/fusion of one or more *(sub)indicators*.

Possible examples of derived indicators are: the fraction nonconforming of a batch received from a supplier, the overall equipment ratio (OEE) of a manufacturing operation, the return of investment (ROI) of a business operation, the Hirsch-index (h) of a scientist, etc.

We remark that the adjective "derived" is often replaced with "aggregated", "composite", "synthetic", "fused", "combined", etc. (Joint Research Centre-European Commission 2008).

3.6 The Condition of "Non-uniqueness" for Indicators

A given *representation target* can be represented by different (independent) indicators or sets of indicators. In other words, the choice of indicators for a certain representation target is not necessary unique. For indicators, the concept of *non-uniqueness* is threefold, as it can be applied to (1) the way of aggregating different indicators into a derived indicator, (2) the choice of single indicators, and (3) the choice of sets of indicators, as illustrated in the following subsections.

3.6.1 "Non-uniqueness" for Derived Indicators

Let us consider an automotive production plant of exhaust-systems with four equivalent production lines: α, β, γ, and δ (Franceschini and Galetto 2004; Franceschini et al. 2007). Figure 3.7 shows a scheme of a generic automotive exhaust system.

In this case, the *context* is the "manufacturing plant" and the *representation target* is the "identification of the best production line on a certain day".

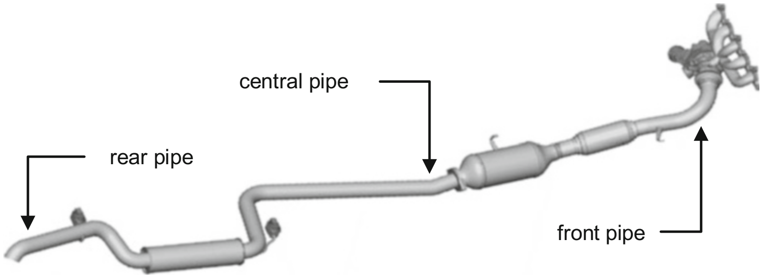

Fig. 3.7 Scheme of an automotive exhaust-system

Table 3.6 Experimental data concerning four equivalent production lines for exhaust-systems, in a manufacturing plant

	Production lines			
Indicators	α	β	γ	δ
Daily production (no. per day)	360	362	359	358
Daily defectiveness (no. per day)	35	32	36	40
Unavailability equipment ratio (%)	4.00	5.50	4.50	5.00

Production line performance is represented by the following three (sub) indicators:

- daily production (the number of items produced in one day);
- daily defectiveness (the number of defective items produced in one day);
- unavailability equipment ratio (the portion of "downtime" in one day).

Let us consider the experimental data reported in Table 3.6.

These three indicators could be aggregated into a unique one. In general, the way of aggregating indicators may depend on several features, such as their scale types, meaning, etc. (Banker et al. 2004).

Weight assignment is a typical solution to reflect the different degree of importance of the individual indicators to be aggregated. Unfortunately, this assignment is often subjective, due to the lack of objective weighing criteria (Franceschini and Maisano 2015; Wang et al. 2015). Of course, the adoption of different weighing criteria may significantly influence/distort the aggregation results (Roberts 1979; Franceschini and Romano 1999). Additionally, the aggregation criteria should be consistent with the scale properties of data and their empirical meaning (Roberts 1979; Franceschini 2001). For example, the adoption of numerical conversion rules (e.g., the use if substitution rates between sub-indicators or specific cost/utility functions) can be often controversial, as it can produce inappropriate distortions in the final result (Fishburn 1973, 1976, 1982; Vincke 1992).

Returning to the problem of interest, we may determine a single ranking of the production lines for each indicator:

daily production:	$\beta \succ \alpha \succ \gamma \succ \delta$
daily defectiveness:	$\beta \succ \alpha \succ \gamma \succ \delta$
unavailability equipment ratio:	$\alpha \succ \gamma \succ \delta \succ \beta$

In this framework, (at least) two different derived indicators (I_B and I_C) can be used to fuse these rankings into a single overall ranking, as explained below.

- **Borda's *indicator*** (I_B)

Borda's indicator is constructed through three basic steps: (i) ranking the alternatives (i.e., production lines) for each (i-th) (sub)indicator (i.e., the three indicators reported in Table 3.6), (ii) associating each alternative (x) with a score corresponding to the its rank position (i.e., 1 for the first position, 2 for the second, and so on), and (iii) synthesizing the scores obtained by each alternative and for each (sub)indicator into an overall score—also known as Borda's score—which is the sum of the rank obtained. The winner is the alternative with the lowest Borda score:

$$I_B(x) = \sum_{i=1}^{m} r_i(x), \tag{3.26}$$

where $r_i(x)$ is the ranking obtained by the line x, considering the i-th indicator, and m is the total number of indicators (in this case $m = 3$).

The winner, i.e., the best line (x^*), is given by (Borda 1781):

$$I_B(x^*) = \min_{x \in S} \{I_B(x)\}, \tag{3.27}$$

where S is the set of alternatives, i.e., the production lines in this case: $S \equiv \{\alpha, \beta, \gamma, \delta\}$.

Applying Borda's method to data in Table 3.2, we obtain the following results:

$$I_B(\alpha) = 2 + 2 + 1 = 5$$

$$I_B(\beta) = 1 + 1 + 4 = 6$$

$$I_B(\gamma) = 3 + 3 + 2 = 8$$

$$I_B(\delta) = 4 + 4 + 3 = 11$$

According to Eq. (3.27), the final ranking is: $\alpha \succ \beta \succ \gamma \succ \delta$.

The winner (i.e. the line with best overall performance) is line α.

- **Condorcet's *indicator*** (I_C)

For each pair of alternatives (i.e., production lines), we determine the number of times that one alternative is ranked higher than one other. In other words, alternative

Table 3.7 Paired-comparison data related to the problem in Table 3.6 (for each paired comparison, it is reported the number of times that one alternative is preferred to the other one) and resulting overall ranking, according to Condorcet's method

	α	β	γ	δ	I_C	Ranking
α	–	1	3	3	1	2nd
β	2	–	2	2	2	1st
γ	0	1	–	3	0	3rd
δ	0	1	0	–	0	3rd

Table 3.8 Empirical data concerning three production lines (1st condition)

	Production lines		
Indicators	α	β	γ
Daily production (no. per day)	367	350	354
Daily defectiveness (no. per day)	35	30	37

x is preferred to alternative y when the number of (sub)indicators for which x exceeds y is larger than the number of (sub)indicators for which y exceeds x. In formal terms, the Condorcet's score of a generic alternative (x) is defined as:

$$I_C(x) = \min_{y \in S - \{x\}} \{i : xPy\}, \tag{3.28}$$

where i is the number of indicators for which x is preferred to y, P is the preference operator, and S is the set of alternatives.

For example, an alternative that is preferred to each of the $(n-1)$ remaining ones will obtain the maximum-possible score of $(n-1)$. The (Condorcet) winner is the alternative x^*, for which:

$$I_C(x^*) = \max_{x \in S} \{I_C(x)\}. \tag{3.29}$$

It can be demonstrated that there cannot be more than one (Condorcet) winner (Condorcet 1785). The application of Condorcet's method to data in Table 3.2 produces the results reported in Table 3.7.

According to Eq. (3.29), the final ranking of production lines is ($\beta \succ \alpha \succ \gamma \approx \delta$); so the best line is β.

Although the Borda's and Condorcet's (derived) indicators (I_B and I_C) are related to the same *representation target*, they may lead to different results.

Regarding the Borda's indicator, it can be demonstrated that it is sensitive to the so-called "*irrelevant alternatives*". According to this assertion, if x precedes y in the final ranking then it is not necessarily given that—when a third relatively "weak" alternative z is added—x still precedes y (Arrow 1963; Fishburn 1970; Nurmi 1987).

Let us exemplify this behaviour considering an exhaust-system production plant with three production lines $\{\alpha, \beta, \gamma\}$, which are compared based on daily defectiveness (see Table 3.8).

Table 3.9 Empirical data concerning three equivalent production lines (2nd condition)

	Production lines		
Indicators	α	β	γ
Daily production (no. per day)	367	350	345
Daily defectiveness (no. per day)	35	30	33

The two corresponding rankings are:
daily production: $\alpha \succ \gamma \succ \beta$
daily defectiveness: $\beta \succ \alpha \succ \gamma$

The resulting Borda scores are:

$$I_B(\alpha) = 1 + 2 = 3$$

$$I_B(\beta) = 3 + 1 = 4$$

$$I_B(\gamma) = 2 + 3 = 5$$

According to Borda's *indicator* (Eq. 3.27), the best line is α.

Now suppose that γ varies its position in the rankings (daily production: from 354 items to 345 items, daily defectiveness: from 37 items to 33 items), while the relative rank positions of α and β do not change (see Table 3.9). The new corresponding rankings are:
daily production: $\alpha \succ \beta \succ \gamma$
daily defectiveness: $\beta \succ \gamma \succ \alpha$

The new resulting Borda scores are:

$$I_B(\alpha) = 1 + 3 = 4$$

$$I_B(\beta) = 2 + 1 = 3$$

$$I_B(\gamma) = 3 + 2 = 5$$

In this case, the best line is β, even if the relative performance of α and β has not changed. This demonstrates the sensitivity to "irrelevant alternatives" (i.e., line γ in this case) of the Borda's indicator. In fact, although the third alternative (γ) plays a marginal role in both conditions, it can determine a significant variation in the final result. Of course, this is not a desirable property.

On the other hand, it can be shown that the Condorcet's method does not necessarily fulfil the property of "transitivity" between alternatives (Fishburn 1977). For example, let us consider an exhaust-system production plant with three lines $\{\alpha, \beta, \gamma\}$, which can be compared based on (i) daily production, (ii) daily defectiveness, and (iii) unavailability equipment ratio (see Table 3.10).

Table 3.10 Empirical data concerning three production lines (3rd condition)

	Production lines		
Indicators	α	β	γ
Daily production (no. per day)	365	362	359
Daily defectiveness (no. per day)	35	32	34
Unavailability equipment ratio (%)	5.50	6.00	4.50

Table 3.11 Paired-comparison data related to the problem in Table 3.10 (for each paired comparison, it is reported the number of times that one alternative is preferred to the other one) and resulting overall ranking, according to Condorcet's method

	α	β	γ	I_C	Ranking
α	–	2	1	1	1st
β	1	–	2	1	1st
γ	2	1	–	1	1st

The three corresponding rankings are:
daily production: $\alpha \succ \beta \succ \gamma$
daily defectiveness: $\beta \succ \gamma \succ \alpha$
unavailability equipment ratio: $\gamma \succ \alpha \succ \beta$

Condorcet's method produces the results reported in Table 3.11.

In this case, the transitivity property is not satisfied and the three production lines are tied. In fact, considering the three possible paired-comparison relations:

$$\alpha \succ \beta; \quad \beta \succ \gamma; \quad \gamma \succ \alpha.$$

I_B and I_C can be considered as independent *indicators*, as there is no mathematical transformation which can univocally link them (Fishburn 1970, 1977).

The results obtained by applying Borda's and Condorcet's *indicators* are different, although the *representation target* is the same. We can deduce that, for a given representation target, the representation by indicators is not unique. Also, the way (sub)indicators are aggregated into *derived* indicators is not necessarily *unique* (see Fig. 3.8).

3.6.2 "Non-uniqueness" for Basic Indicators

Going back to the four production lines $\{\alpha, \beta, \gamma, \delta\}$, suppose that the representation target is "*identifying the line with the lowest defectiveness*". This representation target can be operationalized adopting (at least) two different indicators:

- The indicator "*number of rejected parts*" (I_R), which corresponds to the number of rejected parts in one-day production. The best line (x^*) is the one for which:

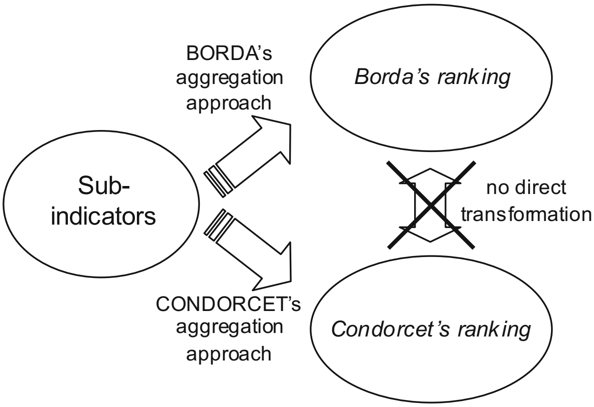

Fig. 3.8 Schematic representation of the "independence" between Borda's indicator and Condorcet's indicator

Table 3.12 Daily defectiveness indicators of four production lines $\{\alpha, \beta, \gamma, \delta\}$. I_R is the number of rejected parts while I_D is the number of defects detected

	I_R	I_D
α	35	43
β	25	39
γ	17	45
δ	21	25

$$I_R(x^*) = \min_{x \in A} \{I_R(x)\} \tag{3.30}$$

being $A \equiv \{\alpha, \beta, \gamma, \delta\}$.

- The indicator "*number of defects detected*" (I_D), which corresponds to the number of defects detected through a 100% control of daily production. We remark that a *defect* is the manifestation of a specific non-conformity while a *defective* is a product unit including one or more defects (Montgomery 2012). Also, the presence of minor defects does not necessarily entail the rejection of a product unit.

  Going back to the point, the better line is the one for which:

$$I_D(x^*) = \min_{x \in A} \{I_D(x)\} \tag{3.31}$$

Again, there is no mathematical transformation which univocally links I_R and I_D. So, this example shows that a certain representation target can be represented by different alternative indicators. In other terms, the choice of *single* indicators is not necessarily unique.

Developing the example, let us consider the daily defectiveness indicators in Table 3.12.

Considering the indicator I_R, the best line is γ:

$$I_R(x^*) = \min_{x \in A} \{35, 25, 17, 21\} = 17 \quad (3.32)$$

On the other hand, considering the indicator I_D, the best line is δ:

$$I_D(x^*) = \min_{x \in A} \{43, 39, 45, 25\} = 25 \quad (3.33)$$

It can be noticed that the use of different indicators can produce different results. While I_R focuses on minimising the number of rejected parts, I_D focuses on optimizing the process phases and reducing every sort of defect. Even though these indicators represent the same representation target, they may affect different actions and decisions. Section 5.8 will go into this point.

3.6.3 "Non-uniqueness" for Measurements

The non-fulfilment of the uniqueness condition may have several consequences. The most evident one is that two or more alternative indicators can be associated with a certain representation target and there will not necessarily be any mathematical transformation linking them.

We remark that even *measurements* do not fulfil the condition of uniqueness (Roberts 1979); in fact, a given measurement is associated with a "*class of equivalent scales*". All the possible transformations between one scale and one other form the so-called "*class of admissible transformations*" (see Sect. 3.2) (Finkelstein 2003).

Regarding measurements, the unclear definition of the measurement process, the incorrect determination of empirical observations/laws, and the non-fulfilment of the condition of *uniqueness* are included into the concept of *uncertainty* (JCGM 100:2008 2008; Finkelstein 2003). A similar concept can be defined for *indicators*.

3.6.4 "Non-uniqueness" when Specializing the Representation Target

A deeper specialization of the representation target does not necessarily eliminate the condition of *non-uniqueness* for indicators.

Let us consider again the set of four lines $A \equiv \{\alpha, \beta, \gamma, \delta\}$ in Table 3.6. We showed that the representation target "identification of the line with lowest defectiveness" can be operationalized by at least two different indicators.

Now, let us further specialize the representation target, trying to avoid the condition of *non-uniqueness* of *indicators*. A more specialized definition may be "identifying the line with the lowest number of rejected parts". To this purpose, at least two different *indicators* can be adopted:

- The indicator "daily number of rejected parts that cannot be reworked" ($I_{R,nr}$). According to this indicator, the best line (x*) is the one for which:

$$I_{R,nr}(x^*) = \min_{x \in A} \{I_{R,nr}(x)\}. \tag{3.34}$$

- The indicator "daily number of rejected parts with two or more defects" ($I_{R,dd}$). According to this indicator, the best line (x*) is the one for which:

$$I_{R,dd}(x^*) = \min_{x \in A} \{I_{R,dd}(x)\}. \tag{3.35}$$

In general, a semantic specialization of a representation target encourages a more accurate definition of the related indicator (or set of indicators), never reaching the condition of *uniqueness*. In fact, although the representation target can be further specialized (e.g., "identifying the line with the lowest average number of rejected parts, with at least two defects"), different alternative options in defining indicators will remain (e.g., a possible indicator could not consider the parts produced during a process slowdown/interruption, while another could include them).

The result is that, for a given representation target, there cannot be a univocal definition of the relevant indicator.

In conclusion, the condition of uniqueness can be considered from two different perspectives. The first one concerns the definition of the representation target (i.e., "what" to represent) and the second one concerns the way to perform the representation (i.e., "how" to represent); the semantic specialization of the *representation target* may refine the first aspect only.

3.6.5 "Non-uniqueness" for Sets of Indicators

A direct consequence of *non-uniqueness* condition is that a representation target can be described by different sets of indicators. This stimulates the need for rules (or empirical procedures) to identify the better one(s). This in turn requires the analysis of the possible *impact* of indicators: in fact, different sets of indicators may differently influence the overall behaviour of a system, with unexpected consequences (Barnetson and Cutright 2000; Hauser and Katz 1998; Franceschini et al. 2013).

According to some authors, when selecting indicators two different families of properties should be considered, such as:

- *basic properties*, directly related to the mathematical definition of indicators (e.g., uniqueness, exhaustiveness, non redundancy, monotony, etc.) (Roberts 1979; Roy and Bouyssou 1993);

- *operational properties*, related to their practical implementation (e.g., validity, robustness, simplicity, integration, economy, compatibility, etc.) (Caplice and Sheffi 1994).

Chapter 4 will discuss in detail some of these properties and other important properties of indicators.

References

Arrow, K. (1963). *Social choice and individual values* (2nd ed.). New Haven, CT: Yale University Press.

Banker, R. D., Chang, H., Janakiraman, S. N., & Konstants, C. (2004). A balanced scorecard analysis of performance metrics. *European Journal of Operations Research, 154*, 423–436.

Barnetson, B., & Cutright, M. (2000). Performance indicators as conceptual technologies. *Higher Education, 40*, 277–292.

Borda, J. C. (1781). Mémoire sur les élections au scrutin, Comptes Rendus de l'Académie des Sciences. Translated by Alfred de Grazia as Mathematical derivation of an election system. *Isis, 44*, 42–51.

Caplice, C., & Sheffi, Y. (1994). A review and evaluation of logistics metrics. *The International Journal of Logistics Management, 5*(2), 11–28.

Cecconi, P., Franceschini, F., & Galetto, M. (2006). Measurements, evaluations and preferences: A scheme of classification according to the representational theory. *Measurement, 39*(1), 1–11.

Cecconi, P., Franceschini, F., & Galetto, M. (2007). The conceptual link between measurements, evaluations, preferences and indicators, according to the representational theory. *European Journal of Operational Research, 179*(1), 174–185.

Churchman, C. W., & Ratoosh, P. (1960). *Measurement: Definitions and theories*. New York: Wiley.

Condorcet M.J.A.N.C. (Marquis de). (1785). *Essai sur l'application de l'analyse à la probabilité des décisions redues à la pluralité des voix*. Paris: Imprimerie Royale.

Dawes, R. M., & Smith, T. L. (1985). Attitude and Opinion measurement. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 1). New York: Random House.

Finkelstein, L. (1982). Handbook of measurement science. In P. H. Sydenham (Ed.), *Theoretical fundamentals* (Vol. 1). New York: Wiley.

Finkelstein, L. (2003). Widely, strongly and weakly defined measurement. *Measurement, 34*(1), 39–48.

Finkelstein, L., & Leaning, M. (1984). A review of the fundamental concepts of Measurement. In L. Gonella (Ed.), *Proceedings of the workshop on fundamental logical concepts of Measurement*. IMEKO, 1983. Reprinted in: Measurement, Vol. 2(1).

Fishburn, P. (1970). *Utility theory for decision making*. New York: Wiley.

Fishburn, P. (1973). *The theory of social choice*. Princeton, NJ: Princeton University Press.

Fishburn, P. C. (1976). Noncompensatory preferences. *Synthese, 33*, 393–403.

Fishburn, P. C. (1977). Condorcet social choice functions. *SIAM Journal on Applied Mathematics, 33*, 469–489.

Fishburn, P. C. (1982). *The foundations of expected utility*. Dordrecht: D. Reidel.

Franceschini, F. (2001). *Dai Prodotti ai Servizi. Le nuove frontiere per la misura della qualità*. Torino: UTET Libreria.

Franceschini, F., & Galetto, M. (2004). An empirical investigation of learning curve composition laws for Quality Improvement in complex manufacturing plants. *Journal of Manufacturing and Technology Management, 15*(7), 687–699.

Franceschini, F., Galetto, M., & Maisano, D. (2007). *Management by measurement: Designing key indicators and performance measurement systems*. Berlin: Springer Science & Business Media.

Franceschini, F., Galetto, M., & Turina, E. (2013). Techniques for impact evaluation of performance measurement systems. *International Journal of Quality and Reliability Management, 30* (2), 197–220.
Franceschini, F., & Maisano, D. (2015). Checking the consistency of the solution in ordinal semi-democratic decision-making problems. *Omega, 57*, 188–195.
Franceschini, F., & Romano, D. (1999). Control chart for linguistic variables: A method based on the use of linguistic quantifiers. *International Journal of Production Research, 37*(16), 3791–3801.
Hauser, J., & Katz, G. (1998). Metrics: You are what you measure! *European Management Journal, 16*(5), 517–528.
ISO/IEC Guide 99:2007. (2007). International vocabulary of metrology – Basic and general concepts and associated terms (VIM). International Organization for Standardization, Geneva. Retrieved September 2018, from http://www.bipm.org/en/publications/guides/vim.html.
JCGM 100:2008. (2008). *Evaluation of measurement data – Guide to the expression of uncertainty in measurement*. Paris: BIPM.
Joint Research Centre-European Commission. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. OECD Publishing. ISBN: 9789264043459.
Kano, N. (1984). Attractive quality and must-be quality. *Hinshitsu (Quality, The Journal of Japanese Society for Quality Control), 14*, 39–48.
Krantz, D., Luce, R., Suppes, P., & Tversky, A. (1971). *Foundations of measurement* (Vol. 1). New York: Academic Press.
Krantz, D., Luce, R., Suppes, P., & Tversky, A. (1989). *Foundations of measurement* (Vol. 2). New York: Academic Press.
Krantz, D., Luce, R., Suppes, P., & Tversky, A. (1990). *Foundations of measurement* (Vol. 3). New York: Academic Press.
Marcus-Roberts, H. M., & Roberts, F. S. (1987). Meaningless statistics. *Journal of Educational Statistics, 12*(4), 383–394.
Mikulić, J., & Prebežac, D. (2011). A critical review of techniques for classifying quality attributes in the Kano model. *Managing Service Quality: An International Journal, 21*(1), 46–66.
Montgomery, D. C. (2012). *Statistical quality control* (7th ed.). New York: Wiley.
Narayana, C. L. (1977). Graphic positioning scale: An economical instrument for surveys. *Journal of Marketing Research, XIV*, 118–122.
Nunally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
Nurmi, H. (1987). *Comparing voting systems*. Dordrecht: D. Reidel.
Roberts, F. S. (1979). *Measurement theory*. Reading, MA: Addison-Wesley.
Roy, B., & Bouyssou, D. (1993). *Aide Multicritère à la Décision: Méthodes et Cas*. Paris: Economica.
Scott, D., & Suppes, P. (1958). Foundational aspects of theories of measurement. *Journal of Symbolic Logic, 23*, 113–118.
Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*(2684), 677–680.
Stevens, S. S. (1959). Measurement, psychophysics, and utility. In C. W. Churchman & P. Ratoosh (Eds.), *Measurement: Definitions and theories* (pp. 18–63). New York: Wiley.
Sydenham, P. H., Hancock, N. H., & Thorn, R. (1989). *Introduction to measurement science and engineering*. New York: Wiley.
Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.
Vincke, P. (1992). *Multi-criteria decision aid*. New York: Wiley.
Wang, B., Liang, J., & Qian, Y. (2015). Determining decision makers' weights in group ranking: A granular computing method. *International Journal of Machine Learning and Cybernetics, 6*(3), 511–521.

Properties of Indicators

4

Abstract

Indicators are helpful tools to represent (complex) processes, supporting evaluations and decisions. Unfortunately, selecting "good" indicators is not so trivial for at least two reasons: (i) there are not organic methods to support this activity and (ii) the success of this activity may depend on the complexity of the process of interest and the experience/intuition of users. The aim of the present chapter is to provide a taxonomy of some desirable properties of indicators, trying to answer several research questions, such as: "How many indicators should be used for representing a certain process?", "Is there an optimal set of indicators?", and "Can we aggregate/fuse multiple (sub)indicators into a single one?". Description is accompanied by many pedagogical examples and an operational procedure to support the construction of indicators.

4.1 Introduction

The previous chapters pointed out that indicators are helpful tools to gather information on complex processes and analyse their evolution. According to what explained in Chaps. 1 and 3, a "set of indicators" is supposed to represent the important dimensions of a process of interest, i.e., it is a "representation model" to support evaluations and decisions on the process itself (Keeney and Raiffa 1976). According Roy and Bouyssou (1993), each model is a "*representation of many phenomena, which are extracted from their context to support the analysis*".[1]

In the current scientific literature, there are many examples of process modeling by means of performance indicators. For example, when dealing with a

[1]This definition can be extended to all the categories of models; for instance, *physical*, *engineering*, *theoretical*, and *empirical* ones.

F. Franceschini et al., *Designing Performance Measurement Systems*, Management for Professionals, https://doi.org/10.1007/978-3-030-01192-5_4

manufacturing plant, indicators like “throughput”, “defectiveness”, “output variability” and “efficiency” are commonly adopted (see Chap. 1) (Brown 1996; Kaydos 1999; Galbraith et al. 1991; Maskell 1991; Smith 2000). Special emphasis is also given to the so called “design metrics”, that is to say those factors in product design that may affect one or more stages of the product lifecycle (Galbraith and Greene 1995).

Historically, logistics and manufacturing functions are some of the factory functions that are mostly concerned with the use of performance indicators (Neely et al. 1995; New and Szwejczewski 1995; Gunasekaran and Kobu 2007). An interesting survey regarding “logistic metrics” is presented by Caplice and Sheffi (1994, 1995). Being convinced that a strategically well-designed performance measurement system can be defective at the individual metric level, these authors state that there is no need for developing new performance metrics (at least in the field of logistics) but there is a lack of methods to evaluate them.

In public-administration management, the concept of “performance measurement” is far from new, as Perrin (1998) states in his literature review. Performance measurements have been widely promoted by governments for several decades, for the purpose of increasing management’s focus on achieving results (Winston 1999). This is further demonstrated by the publication of “The Performance-Based Management Handbook” by Oak Ridge Institute for Science and Education (ORISE) (U.S. Department of Energy – PBM SIG 2012).

Many authors have tried to address their studies towards the definition of practical guidelines to assist practitioners in the definition of performance measurement systems (Caplice and Sheffi 1994; Hauser and Katz 1998; Lohman et al. 2004; Gunasekaran and Kobu 2007). Nevertheless, a general and organic theory to support the selection of indicators for the representation of a generic process is still missing.

Process modeling by means of indicators raises several questions: “How many indicators shall we use?”, “Is there an optimal set?”, “Is this set unique?”, “If not, how can we identify the best one (assuming that it exists)?”, “Can multiple indicators be aggregated into a unique one?”, etc.

Selecting good indicators is not so trivial for (at least) two reasons: (i) there are not organic methods to support this activity and (ii) the success of this activity may depend on the complexity of the process of interest and the experience/intuition of users. Chapter 5 will discuss some (relatively) diffused methods to do it.

The aim of the present chapter is to provide a taxonomy of some desirable properties of indicators, trying to answer the above questions. Description is accompanied by a large number of pedagogical examples. Furthermore, an operational procedure to support the construction of indicators will be illustrated.

4.2 Single and Aggregated Performance

When defining indicators, two performance levels should be considered:

- *Single (or local) level*. Each indicator is focused on the representation of a specific dimension of the process of interest.
- *Aggregated (or global) level*. The relevant dimensions of the process (or a portion of it) are represented by a set of indicators, which are often aggregated into other (derived) indicators.

Example 4.1 In a university entrance exam, the best 200 students are selected according to the following three indicators:

I_1 "grade of the high school qualification";
I_2 "result of a written entrance test";
I_3 "result of an oral entrance test".

Each of these indicators is expressed by a number included between 0 and 100. The overall performance is calculated considering a weighted average: weights of the three indicators (I_1, I_2, I_3) are respectively 5/10, 3/10 and 2/10. The resulting (derived) indicator (I_{TOT}) is:

$$I_{TOT} = \frac{5}{10} \cdot I_1 + \frac{3}{10} \cdot I_2 + \frac{2}{10} \cdot I_3. \tag{4.1}$$

It can be noticed that the high school qualification grade (I_1) has the largest weight, while the result of the oral entrance test (I_3) has the lowest one. Although the choice of weights can be questionable, it may have important effects on the overall-performance indicator (I_{TOT}). Different weights may determine different strategies of selection of students.

4.3 The Concept of Indicator

Recalling what is formalized in Sect. 3.4.2, an indicator is a not-necessarily-objective homomorphic mapping of a set of empirical manifestations into a set of symbolic manifestations (see Fig. 4.1).

In addition, this mapping is not necessarily one-to-one: manifestations that are undistinguished from the perspective of the representation target are mapped into the same symbol (e.g., in Fig. 4.1, manifestations a_1 and a_2 are undistinguished and therefore mapped into the same symbol z_1).

Chapter 3 has extensively discussed the condition of *non-uniqueness* of the representation. The scheme in Fig. 4.2 summarizes this concept for single indicators (cf. Sect. 3.5.2).

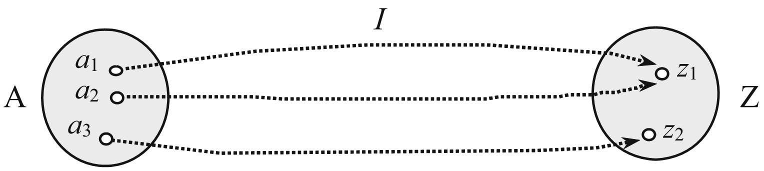

Fig. 4.1 Schematic representation of the concept of indicator. An indicator (I) is defined as a homomorphism that maps a set (A) of empirical manifestations into a set (Z) of symbolic manifestations. In formal terms I: A→Z

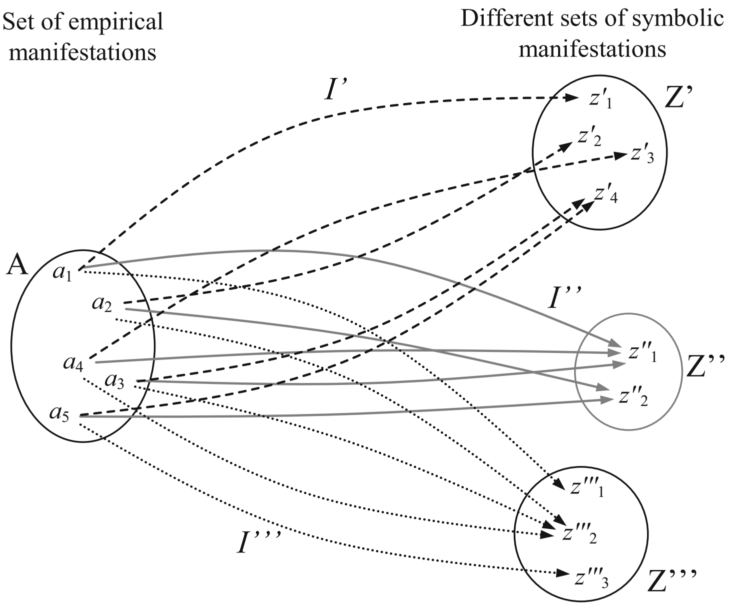

Fig. 4.2 Schematic representation of the condition of *non-uniqueness* for single indicators. The same representation target is operationalized by different indicators (*I'*, *I"* and *I"'*, in this case). Some empirical manifestations, which are undistinguished by one indicator, can be distinguished by another one (e.g., manifestations a_3 and a_5, which are undistinguished by *I'*, are distinguished by *I"* and *I"'*)

Since different indicators may operationalize the same representation target, some questions arise: "What is the best way of selecting them?"; "Is the representation exhaustive?".

A tool that algorithmically generates a set of indicators for a given representation target is not actually conceivable (Roy and Bouyssou 1993). Such a tool would have to automate several conceptual/semantic operations: (1) definition of a set of indicators, (2) testing/verification, (3) correction of some indicators, validation, etc.

Example 4.2 In a manufacturing company producing hydraulic valves, the following representation target is identified: "improving the quality of the production output". The following set of indicators is implemented to operationalize the representation target:

I_1 "number of units produced in a specific time window";
I_2 "number of units classified as defective and then rejected".

An alternative set of indicators is defined as:
I'_1 "number of units produced by the first of four production lines available";
I'_2 "average percentage of defective units, resulting from a spot inspection of a portion (e.g., 5%) of the total production".

Even though there are many ways of representing a process, identifying the best one can be problematic. The rest of this chapter will describe some (desirable) properties to support the selection/aggregation of indicators and their verification.

As shown in Sect. 3.4.2, there is a relevant link between the concept of *measurement* and that of *indicator*; in fact, measurements represent a subset of indicators. The following examples emphasize this aspect.

Example 4.3 The wheelbase of a motor vehicle (i.e., the geometrical distance between two car-axles) is a dimensional measurement and therefore an indicator too. In fact, the (i) homomorphical mapping of empirical manifestations into numbers is objective and (ii) the relations among numbers isomorphically reflect the relations among empirical manifestations.

Example 4.4 The representation target "classification of objects depending on their size" is operationalized by the indicator "volume of the object" (expressed in cm^3), which is also a measurement. Relations of *(in)equality* (e.g., volume A is different from volume B), *order* (e.g., volume A is lager/lower than volume B), *interval* (e.g., the difference between volume A and volume B is larger than x cm^3), and *ratio* (e.g., volume A is x times larger than volume B), which are included in the empirical relational system, are mapped into corresponding mathematical relations in the symbolic relational system (Franceschini 2001; Finkelstein 2003).

Example 4.5 The indicator "comfort of a car seat for passengers", which can be expressed on an ordinal scale from 1 to 5 (e.g., 1-very uncomfortable, 2-uncomfortable, 3-neutral, 4-comfortable, 5-very comfortable), is not a measurement, due to lack of objectivity.

Example 4.6 Let us consider the representation target "classification of the students of a class", which is operationalized by the indicator "name of a student". This indicator links one student (empirical manifestation) with his/her name (symbolic manifestation). In nature, there is no order relation among empirical manifestations (i.e., students), which corresponds to the alphabetical (order) relation among the symbolic manifestations (i.e., names). Due to the missing isomorphism between empirical relations and symbolic relations, the "name of a student" is not a measurement.

4.4 Further Considerations on the Classification of Indicators

Any process can be represented through indicators. Usually, the more complex the process, the larger the number and the variety of indicators to be used (Melnyk et al. 2004). A seen in Chap. 3, indicators can be classified into *objective* and *subjective* (Franceschini et al. 2006b).

Consistently with Definition 3.9, an indicator is *objective* when the mapping of empirical manifestations into symbolic manifestations is objective (i.e., it does not change from subject to subject). Otherwise, it is *subjective*. The following examples clarify this concept.

Example 4.7 Let us consider the indicator: "amount of production of a plant". The empirical manifestations (production output) are objectively connected to corresponding symbolic manifestations (number of units). In fact, excluding any counting mistake, different people (or even automatic devices) will obtain the same result.

Example 4.8 The two indicators (i) "preference of a voter at elections" and (ii) "evaluation of the aesthetics of a car" are both subjective, since the mapping of empirical manifestations into symbolic manifestations may depend on subjective perceptions or opinions.

Example 4.9 The indicator "exterior style of a car" is based on the evaluation of an expert using a five-level ordinal scale (1-very bad; 2-poor; 3-fair; 4-good; 5-excellent). The indicator is *subjective* because the same empirical manifestation (i.e., specific car design) can be mapped to different symbolic manifestations (i.e., the five-scale levels), depending on the subject (i.e., the expert).

Objectivity is a *sine-qua-non* condition for measurements. However, objective indicators are not necessarily measurements. Returning to Example 4.6, we notice that—although the "name of a student" is not a measurement—it is certainly an objective entity (see also the discussion of the concept of *conventional objectivity*, in Sect. 5.8).

As explained in Sect. 3.5.2, indicators can also be classified in two more categories:

- *Basic* indicators. They are obtained from a direct observation of an empirical system.
- *Derived* indicators. They are obtained combining the information of one or more (sub)indicators (basic or derived).

Example 4.10 Let consider the derived indicator "percentage of defectives in a production line" (I_3), which is given by the aggregation of two sub-indicators:

$$I_3 = \frac{I_1}{I_2}, \tag{4.2}$$

being:

I_1 number of defective units;
I_2 total number of manufactured units.

Example 4.11 An organization for environmental protection commissions two local Agencies (A and B) to estimate the pollution level of the exhaust of motor vehicles, on the basis of four pollutants:

I_{NO_X}: the concentration of NO_X [μg/m^3];
I_{HC}: the concentration of unburnt hydrocarbons [μg/m^3];
I_{CO}: the concentration of CO [μg/m^3];
$I_{PM_{10}}$: the concentration of PM_{10} [μg/m^3].

Agency A maps each concentration into a three-level ordinal scale (1-harmless; 2-acceptable; 3-unacceptable for human health) and specifies four corresponding derived indicators ($I'_{NO_X}, I'_{HC}, I'_{CO}, I'_{PM_{10}}$). The latter indicators are then aggregated into the following derived indicator (see also the discussion on air quality indicators in Chap. 2):

$$I^A_{TOT} = \max\left\{I'_{NO_X}, I'_{HC}, I'_{CO}, I'_{PM_{10}}\right\}. \tag{4.3}$$

Figure 4.3 depicts the aggregation mechanism.

Agency B maps the concentration of each pollutant into a five-level ordinal scale, defining four corresponding derived indicators ($I''_{NO_X}, I''_{HC}, I''_{CO}, I''_{PM_{10}}$). Then the latter indicators are aggregated into another derived indicator:

$$I^B_{TOT} = \frac{\left(I''_{NO_X} + I''_{HC} + I''_{CO} + I''_{PM_{10}}\right)}{4}. \tag{4.4}$$

While Agency A estimates the pollution level on the basis of the predominant pollutant(s), Agency B estimates it on the basis of their average concentration.

Example 4.11 shows that, for a given representation target, the same (sub) indicators can be aggregated in different ways. A detailed description of popular ways of aggregating air quality indicators is contained in Sect. 2.3. Additionally, the example shows that *derived* indicators may (in turn) be aggregated into a *derived* indicator of higher grade. Extending this concept to the limit, the totality of indicators in use could recursively be aggregated into a single "super-indicator", which depicts the global performance (see Fig. 4.4). This is indeed a very challenging activity, even though it may inevitably lead to debatable (over)simplifications

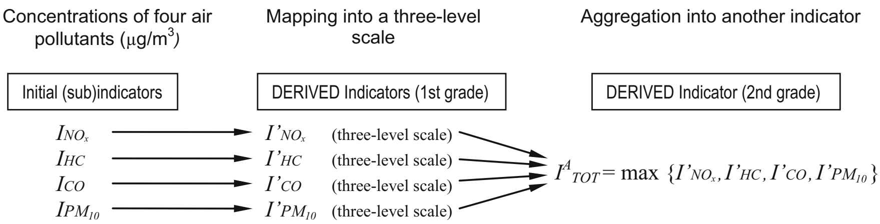

Fig. 4.3 Scheme of the construction and aggregation of air-quality *derived* indicators by Agency A

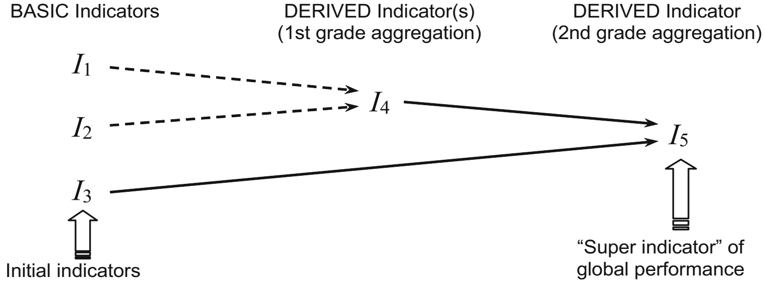

Fig. 4.4 Representation of the concept of *global performance*. The initial (*basic*) indicators are (recursively) aggregated into a single "super indicator"

(Melnyk et al. 2004). Chapter 5 exhaustively discusses to what extent this recursive aggregation process is reasonable.

Returning to Example 4.11, we note that when aggregating the four (sub) indicators I''_{NO_X}, I''_{HC}, I''_{CO} and $I''_{PM_{10}}$ into I^B_{TOT}, they are treated as if they were defined on a *ratio* scale—i.e., a scale with non-arbitrary zero and meaningful distance—even if they actually are not. In fact, these ordinal-scale values are improperly "promoted" to ratio-scale values, in the moment in which they are combined through (weighted) averages or—more in general—some statistics that are permissible to ratio-scale values only (Stevens 1946; Roberts 1979; Franceschini 2001).

4.4.1 The Representational Approach for Derived Indicators

The concept of *derived indicator* can also be interpreted according to the *representation theory of indicator*. The empirical system of a derived indicator is replaced with the combination of the symbolic manifestations of sub-indicators (e.g., through the Cartesian product). This combination is then homomorphically mapped into further symbolic manifestations (see Fig. 4.5).

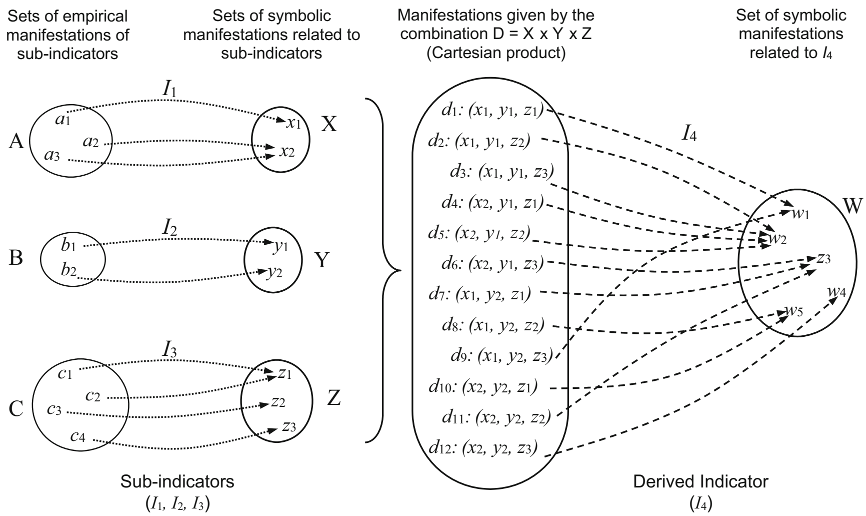

Fig. 4.5 Schematic representation of a derived indicator, according to the *representation theory of indicator*

The aggregation of several (sub)indicators into a derived one can be "tricky", especially for complex processes. Let us consider the following example.

Example 4.12 To investigate the "health state of a manufacturing company", different indicators are adopted, such as "net profit", "throughput", "market share", "customer satisfaction", etc. The aggregation of these (heterogeneous) indicators can result in the introduction of questionable simplifications.

4.5 Brief Outline of the Properties of Indicators in Literature

We analyzed the existing scientific literature with the purpose of finding a set of desirable properties of indicators for a suitable process representation. The description of these properties is often unstructured and without any formal mathematical approach.

Table 4.1 reports the major properties of indicators, according to several authors in the literature (Caplice and Sheffi 1994). These properties concern individual indicators exclusively, without any distinction among *basic*, *derived* or *sets* of indicators.

Table 4.1 Properties of individual indicators, according to several authors (Caplice and Sheffi 1994)

Requirements of the indicators	Properties defined by different authors					
	Mock and Groves (1979)	Edwards (1986)	Juran (1988)	NEVEM (1989) AT Kearney (1991)	Mentzer and Konrad (1991)	Caplice and Sheffi (1994)
Does the indicator capture actual events and activities accurately?	Valid	Reliable		Valid		Validity
Does the indicator control for inherent errors in data collection? Is it repeatable?	Reliable				Measurement error	Robustness
Is the indicator using a correct type of numerical scale for the values?	Scale type					Behaviourally sound
Is the indicator still accurate if the scale is transformed to another type?	Meaningful					Behaviourally sound
Do the benefits outweigh the costs of using the indicator?	Economical worth	Cost/ benefit	Economical	Profitability		Economy
Will the indicator create incentives for improper or counterintuitive acts?	Behavioural implications				Human Behaviour	Behaviourally Sound
Does the indicator use data currently available from the existing ones?		Available				Compatibility
Is the indicator compatible with the existing information system and flow?			Compatible to existing systems	Compatible		Compatibility
Does the indicator provide a guide for an action to be taken?		Useful		Utility		Usefulness
Can the indicator be compared across time, location, and organizations?		Consistent	Apply broadly	Comparable	Comparable	Robustness
Is the indicator viewed and interpreted similarly by all affected parties?			Uniform interpretation & agreed upon basis			Robustness
Is the indicator simple and straightforward enough to be easily understood by those affected?			Understandable			Usefulness
Does the indicator include and measure all of the important components and aspects of the system?				Covering potential	Under-determination	Integration
Does the indicator promote coordination across the various players in the supply chain?						Integration
Is the indicator of a sufficient level of detail and precision for a decision maker?				Accurate		Level of detail

With permission

4.6 Taxonomy of the Properties of Indicators

Table 4.2 contains a taxonomy of (desirable) properties of indicators, based on the one proposed in Franceschini et al. (2008). Properties are classified into four groups: properties of *single* indicators, properties of *sets* of indicators, properties of *derived* indicators, and *accessory* properties. The evolution of these properties may support the selection of performance indicators in different contexts.

It can be reasonably assumed that a large part of the properties found in the literature (see Table 4.1), can be incorporated in this taxonomy. Section 4.6.1 will illustrate the properties contained in Table 4.2, with the support of a number of pedagogical examples.

Table 4.2 Taxonomy of the properties of indicators, based on the classification described in Sect. 4.4

Category	Properties	Short description
Properties of single indicators	Consistency with representation target	The indicator should properly represent the representation target of interest.
	Meaningfulness of statements	The meaningfulness of statements involving a generic indicator is certainly a desirable condition.
	Level of detail	The indicator should not provide more information than necessary.
	Counter-productivity	Indicators should not encourage counter-productive actions.
	Economic impact	Data collection and elaboration should be economically sustainable.
	Simplicity of use	The indicator should be easy to understand and use.
Properties of sets of indicators	Exhaustiveness	Indicators should cover the important dimensions of the process and represent them in a "balanced" way.
	Non-redundancy	Every set should not include redundant indicators.
Properties of derived indicators	Monotony	The derived indicator should "respond" to variations in one or more sub-indicators.
	Compensation	Variations in individual sub-indicators can compensate with each other, without producing any variation in the derived indicator.
Accessory properties	Long-term goals	Representation targets should be consistent with the long-term goals of the organization of interest.
	Customer orientation	Representation targets should be customer oriented.

4.6.1 Properties of Single Indicators

The following properties refer to single indicators (both basic and derived).

Consistency with Representation Target

As seen in Sect. 3.3, one indicator should properly operationalize a representation target. Although this notion may seem pleonastic, indicators are often implemented just because they are familiar to users and/or easily obtainable, not necessarily because they better operationalize representation target(s) (Franceschini and Maisano 2017).

Consistency between representation targets and indicators should be thoroughly verified (Denton 2005). This concept is exemplified below.

Example 4.13 Referring to the representation target "sales of a manufacturing company", the indicator I_s "total number of goods sold annually" is used to represent the process. Later, company managers realize that *quarterly* information on sales would be more useful for estimating seasonal trends. Consequently, the new indicator I'_s "total number of goods sold quarterly" can replace the first one.

According to the representation target, the second indicator is more accurate than the first one, as it includes some important empirical manifestations (i.e., quarterly information on sales) that are ignored by I_s.

Section 5.2.1 will illustrate a practical methodology to support the definition of indicators that are consistent with representation targets.

Meaningfulness of Statements

Section 3.2.6 debated the meaningfulness of statements concerning measurements, pointing out that the correct use of a measurement scale depends on the statements regarding the objects under that scale. The same notion can be extended to indicators (both basic and derived).

While we are aware that indicators do not necessarily require the isomorphism F: R → P (see Sect. 3.3), the fact that a statement is meaningful generally implies that the relations among symbolic manifestations (i.e., symbols/numbers) reflect those among empirical manifestations, (i.e., objects) without the latter being unduly promoted or degraded. For this reason, the meaningfulness of statements involving a generic indicator is certainly a desirable condition (Narens 1996, 2002).

The considerations about the meaningfulness/meaninglessness of statements involving measurements can be extended to *basic* indicators. Precisely, Definition 3.5 can be slightly modified into:

Definition 4.1 "A statement involving a *basic* indicator is called meaningful if its truth or falsity is unchanged whenever applying any admissible scale transformation to the indicator itself"; otherwise, it is called meaningless.

With some more substantial changes, the same concept can be extended to *derived* indicators:

Definition 4.2 "A statement involving a *derived* indicator is called meaningful if its truth or falsity is unchanged whenever applying any admissible scale transformation to one or more of its corresponding sub-indicators"; otherwise, it is called meaningless.

It can be noticed that the definition for derived indicators is similar to the one for basic indicators, with the difference that scale transformations should be applied at the level of sub-indicators.

The verification mechanism is similar to that seen in Sect. 3.2.6: i.e., while the meaningfulness of a statement should be demonstrated through a general (analytical) proof, its meaninglessness can be formally demonstrated through either a general (analytical) proof or a single counter-example. In addition, Rules 3.1 and 3.2 (see Sect. 3.2.6) also apply. Here are some application examples.

Example 4.14 A university entrance exam includes two tests (A and B). Each test is evaluated through a score between 0 and 100. For both tests, scores are assumed to be distributed normally with mean $\mu = 50$ and standard deviation $\sigma = 10$. Two possible aggregation approaches to determine an overall score (OS) of candidates are: (1) adding the scores of the individual tests (hereafter denominated A and B); (2) adding the *cumulative probabilities* (or *percentile ranks*) of the corresponding scores.

We now compare and critically analyse the two aggregation approaches, answering the following questions:

(i) Considering two candidates (X and Y) with scores $A_X = 32$, $B_X = 40$ and $A_Y = 35$, $B_Y = 37$ respectively, what are their overall scores (OS_X and OS_Y), according to both the proposed approaches?
(ii) Let us consider a generic (additive) aggregation model $OS_i = I_{A_i} + I_{B_i}$; for which scale type(s) of sub-indicators (I_{A_i} and I_{B_i}), is the statement $OS_X > OS_Y$ meaningful?
(iii) In light of previous results, what is the best of the two aggregation approaches and why?

In the first approach, the initial sub-indicators (i.e., scores in tests A and B) are aggregated by a sum. In the second approach, before being summed, they are transformed through the so-called *cumulative probability density function* related to a normal distribution $N(\mu, \sigma^2)$:

$$F(x) = \int_{-\infty}^{x} f(x) \cdot dx = \int_{-\infty}^{x} \left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\right] \cdot dx = P\left(z < \frac{x-\mu}{\sigma}\right). \quad (4.5)$$

We note that $F(x)$ is a strictly-monotonous increasing non-linear function.

Applying approach (1), the following results are obtained:

$$\left.\begin{array}{l} A_X = 32 \\ B_X = 40 \end{array}\right\} \Rightarrow OS_X = A_X + B_X = 72 \\ \left.\begin{array}{l} A_Y = 35 \\ B_Y = 37 \end{array}\right\} \Rightarrow OS_Y = A_Y + B_Y = 72 \quad . \tag{4.6}$$

Applying approach (2), the following results are obtained:

$$\left.\begin{array}{l} F(A_X) = P\left(z < \dfrac{A_X - \mu}{\sigma}\right) = P(z < -1.8) = 3.59\% \\ F(B_X) = \Pr\left(z < \dfrac{B_X - \mu}{\sigma}\right) = P(z < -1) = 15.87\% \end{array}\right\} \Rightarrow OS_X = F(A_X) + F(B_X) = 19.46\% \\ \left.\begin{array}{l} F(A_Y) = \Pr\left(z < \dfrac{A_Y - \mu}{\sigma}\right) = P(z < -1.5) = 6.68\% \\ F(B_Y) = \Pr\left(z < \dfrac{B_Y - \mu}{\sigma}\right) = P(z < -1.3) = 9.68\% \end{array}\right\} \Rightarrow OS_Y = F(A_Y) + F(B_Y) = 16.36\%. \tag{4.7}$$

We notice that approach (1) produces the same result for both candidates ($OS_X = OS_Y$). On the other hand, according to approach (2), $OS_X > OS_Y$.

The example shows that, in general, the two approaches may lead to different results. To be precise, they would systematically lead to the same results only in the case in which test scores follow a uniform distribution. In fact, since the $F(x)$ related to a uniform distribution is a linear monotonically increasing function like $\Phi(x) = a{\cdot}x + b$, with $a > 0$, it would not distort the "distance" relation between test scores.

Let us now return to the proposed aggregation approaches. Generalizing, *OS* can be seen as the sum of two sub-indicators (I_A and I_B): $I_A = A$ and $I_B = B$ in approach (1), while $I_A = F(A)$ and $I_B = F(B)$ in approach (2).

Ignoring for a moment the real scale types of sub-indicators, we now determine the ones for which the aforementioned statement would be meaningful. Starting from the most "powerful" scale type, we analyse whether the statement would be meaningful for sub-indicators defined on *ratio* scales. Applying the same permissible scale transformation, $\Phi(x) = a{\cdot}x$ ($a > 0$), to the sub-indicators in the statement of interest, we obtain:

$$\begin{array}{l} OS_X^{'} = a \cdot (I_{AX} + I_{AX}) = a \cdot OS_X \\ OS_Y^{'} = a \cdot (I_{AY} + I_{AY}) = a \cdot OS_Y \end{array} . \tag{4.8}$$

It is immediate to note that the transformations introduced do not alter the initial statement:

$$OS_X^{'} > OS_Y^{'} \Rightarrow \alpha \cdot OS_X > \alpha \cdot OS_Y \Rightarrow OS_X > OS_Y. \tag{4.9}$$

This proves that the statement would be meaningful for sub-indicators defined on ratio scales.

Now we hypothesize that I_A and I_B were defined on *interval* scales, which therefore admit scale transformations like $\Phi(x) = a \cdot x + b$, with $a > 0$. By applying this transformation to the sub-indicators involved in the previous statement, we obtain:

$$\begin{array}{l} OS_X^{'} = a \cdot (I_{AX} + I_{AX}) + 2 \cdot b = a \cdot OS_X + 2 \cdot b \\ OS_Y^{'} = a \cdot (I_{AY} + I_{AY}) + 2 \cdot b = a \cdot OS_Y + 2 \cdot b \end{array}. \tag{4.10}$$

Again, the transformation introduced does not alter the initial statement:

$$OS_X^{'} > OS_Y^{'} \Rightarrow a \cdot OS_X + 2 \cdot b > a \cdot OS_Y + 2 \cdot b \Rightarrow OS_X > OS_Y. \tag{4.11}$$

In conclusion, the statement of interest would be meaningful even for sub-indicators defined on interval scales.

Let us now hypothesize that I_A and I_B were defined on *ordinal* scales. According to what seen in Sect. 3.2.6, the sum of objects defined on ordinal scales is prohibited (see Table 3.1). It is not difficult to construct a numeric counter-example for which the initial statement is altered.

The example in Table 4.3 proves that the statement $OS_X > OS_Y$ would be meaningless for sub-indicators (I_A and I_B) defined on ordinal scales and, therefore, even for sub-indicators defined on less powerful scales (i.e., nominal).

Having established that the use of OS is appropriate *exclusively* for sub-indicators defined on interval and ratio scales, we conclude that approach (1) is formally correct: in fact it seems reasonable to assume that test scores (A and B) are proportional to the level of performance achieved in each test, so they can be seen as objects defined on a ratio scale. On the other hand, the approach (2) leads to summing cumulative probabilities, i.e., objects defined on ordinal scales. For the above reasons, we conclude that the approach (2), apart from being more complicated, is conceptually incorrect.

To further clarify the point relating to the (improper) sum of percentile ranks in approach (2), we present a pedagogical example (Stockburger 2016; Franceschini and Maisano 2017). Let us consider the comparison of scores obtained by two high-school students (Suzy and Johnny) in two tests, the first one in English and the second one in Mathematics. If the scores are distributed normally, then percentile

Table 4.3 Numerical example in which the statement $OS_X > OS_Y$ is initially true and then—after applying the (strictly monotonically increasing) transformation $\Phi(x) = x^3$ to sub-indicators I_A and I_B—it becomes false

	(1) Before transform.			(2) After transform.		
	I_A	I_B	OS	I_A'	I_B'	OS'
Candidate X	4	3	7	64	27	91
Candidate Y	5	1	6	125	1	126

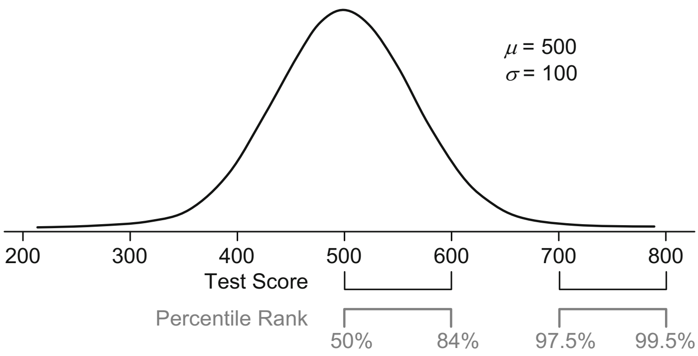

Fig. 4.6 Distribution of the scores obtained in a certain test by a population of high-school students; adapted from Stockburger (2016). With permission

ranks underestimate large differences in the tails of the distribution and overestimate small differences in the middle of the distribution. This is most easily understood in the illustration in Fig. 4.6.

In the above illustration, two standardized achievement tests with $\mu = 500$ and $\sigma = 100$ were given. In the first one, the English test, Suzy made a score of 500 and Johnny made a score of 600, thus there was a 100 point difference between their raw scores. In the second one, the Mathematics test, Suzy made a score of 800 and Johnny made a score of 700, again a 100 point difference in raw scores. It therefore can be said that the differences in the scores on the two tests were equal: 100 points each.

When converted to percentile ranks, however, the differences are no longer equal. In the English test Suzy receives a percentile rank of 50 while Johnny gets an 84: a difference of 34 percentile rank points. On the Mathematics test, Johnny's score is transformed to a percentile rank of 97.5 while Suzy's percentile rank is 99.5: a difference of only two percentile rank points.

It can be seen, then, that a percentile rank has a different meaning depending upon whether it occurs in the middle or the tails of the distribution; differences in the middle of the distribution are magnified, differences in the tails are minimized.

This reasoning can obviously be extended to no-matter-what other (non-uniform) distributions. The lesson learnt from this example is that not only do percentile ranks destroy the interval property but they also destroy the information in a particular manner. Paraphrasing the concept, summing or subtracting percentile ranks is conceptually wrong and misleading.

Example 4.15 FMECA (*Failure Mode, Effects and Criticality Analysis*) is a methodology used in the design phase of products, to identify, analyze and eliminate potential failures, which may deteriorate/compromise the relevant functionalities. This methodology decomposes the product into elementary parts and analyses their so-called *failure modes*. For each *i*-th failure mode, a work team evaluates three factors:

Table 4.4 Example of reference table related to the S_i (severity) scores, for a product FMECA

Score	Description
10	The extent of the failure is such that customer's safety can be at risk
9	The failure causes the deterioration of some primary functions of the product
7–8	The customer is strongly dissatisfied and requires technical assistance
4–6	The customer is dissatisfied but does not require technical assistance
2–3	The customer notices the presence of the failure but tolerates it
1	The customer does not notice even the presence of the failure

(S_i) *severity* related to the possible negative effects;
(O_i) probability of *occurrence*;
(D_i) probability of not detecting the failure (*detectability*).

Each of these factors is associated with a (qualitative) score from 1 to 10, where 1 represents the minimum-risk and 10 the maximum-risk condition; for example, regarding S_i, the higher the score, the greater the negative effects of the failure mode. Scores are assigned by the work team, using some reference tables (see the example in Table 4.4).

FMECA makes it possible to identify the most critical failure modes through a derived indicator: i.e., the *Risk Priority Number* of each *i*-th failure mode, which is defined as $RPN_i = S_i \cdot O_i \cdot D_i$. Precisely, failure modes are sorted in descending order with respect to their RPN_i values; the most critical ones—i.e., those deserving more attention in terms of improvement and design changes—are at the top of the ranking (Franceschini and Galetto 2001).

We now analyse the derived indicator RPN_i, answering the two following questions:

(i) Is the use of this indicator consistent with the scale properties of the relevant sub-indicators S_i, O_i and D_i?
(ii) Is the RPN_i indicator also a measurement?

As anticipated, the RPN_i is used to rank failure modes, which is equivalent to formulating statements between two generic failure modes (*i*-th and *j*-th one) like $RPN_i > RPN_j$. Ignoring for a moment the real scale types (i.e., nominal, ordinal, interval or ratio) of sub-indicators, we now try to understand for which ones the aforementioned statements would be meaningful.

Starting from the more "powerful" scale types, we analyse whether the statement $RPN_i > RPN_j$ (i.e., $S_i \cdot O_i \cdot D_i > S_j \cdot O_j \cdot D_j$) would be meaningful, when assuming that sub-indicators were defined on *ratio* scales. Ratio scales allow similarity-type transformations:

$$\begin{aligned} S_i' &= \Phi(S_i) = a \cdot S_i \quad a > 0 \\ O_i' &= \Phi(O_i) = b \cdot O_i \quad b > 0. \\ D_i' &= \Phi(D_i) = c \cdot D_i \quad c > 0 \end{aligned} \tag{4.12}$$

By applying the scale transformations to sub-indicators, we obtain:

$$\begin{aligned} RPN_i' = S_i' \cdot O_i' \cdot D_i' = (a \cdot b \cdot c) \cdot (S_i \cdot O_i \cdot D_i) = (a \cdot b \cdot c) \cdot RPN_i \\ RPN_j' = S_j' \cdot O_j' \cdot D_j' = (a \cdot b \cdot c) \cdot \left(S_j \cdot O_j \cdot D_j\right) = (a \cdot b \cdot c) \cdot RPN_j \end{aligned}. \tag{4.13}$$

The transformations introduced do not alter the statement, in fact:

$$\begin{aligned} RPN_i' > RPN_i' \;&\Rightarrow\; (a \cdot b \cdot c) \cdot RPN_i > (a \cdot b \cdot c) \cdot RPN_j \\ &\Rightarrow\; RPN_i > RPN_j. \end{aligned} \tag{4.14}$$

This proves that the statement of interest would be meaningful for sub-indicators defined on ratio scales.

Now we hypothesize that S_i, O_i and D_i were defined on *interval* scales, which admit (positive) linear transformations like: $X' = \Phi(X) = a{\cdot}X + b$, with $a > 0$. According to Stevens' *theory of scales of measurement*, the product of objects defined on interval scales is prohibited (see Sect. 3.2); it is not difficult to construct a numerical counter-example showing that the statement would be meaningless in this case.

Let us consider the example in Table 4.5a, where the statement is initially false ($RPN_i < RPN_j$). Applying the scale transformation $X' = X + 2$ ($a = 1$ and $b = 2$) to the sub-indicators S_i and S_j, the results reported in Table 4.5b are obtained, showing that the statement has become true ($RPN_i > RPN_j$). The same considerations can be extended to the other sub-indicators (O_i and D_i).

This proves that the statement of interest would be meaningless for sub-indicators defined on interval scales and, consequently, even for sub-indicators defined on less powerful scales (i.e., ordinal and nominal).

Returning to the definition of S_i, O_i and D_i, we notice that these sub-indicators are defined on ten-level *ordinal* scales; in addition, they are *not objective*, as the assignment of scores, even if guided by dedicated reference tables, is a (at least partly) subjective operation. Due to this lack of objectivity, the RPN_i indicator cannot be considered a measurement.

We can conclude that the typical use of RPN_i is not consistent with the real scale properties of S_i, O_i and D_i, which are actually promoted from (subjective) *ordinal* scales to *ratio* scales. With this, we do not want to discredit such a widespread and

Table 4.5 Numerical example in which the statement $RPN_i > RPN_j$ is initially false but, after the application of the transformation $X' = X + 2$ to S_i and S_j, it becomes true

	(a) Before transform.				(b) After transform.			
	S	*O*	*D*	*RPN*	*S'*	*O*	*D*	*RPN'*
i-th failure mode	3	2	3	18	5	2	3	30
j-th failure mode	7	1	3	21	9	1	3	27

well-established procedure as FMECA but just highlight that FMECA includes questionable aggregations (Franceschini and Galetto 2001).

In general, the aggregation of sub-indicators—that are not defined on a *ratio* scale—through a multiplicative model violates the permissible relations among objects (see also Table 3.1).

Level of Detail

The level of detail is an important aspect to consider when designing an indicator. One indicator with exaggerated level of detail provides more information than necessary, so it could complicate the analysis and be economically wasteful. Reversing the perspective, if (i) one indicator maps two empirical manifestations into different symbolic manifestations and (ii) these empirical manifestations are indistinguishable from the perspective of the representation target, then the mapping resolution will be higher than necessary and the *level of detail* will be exaggerated (see Fig. 4.7a). This condition can be checked analysing the representation target carefully.

In formal terms:

<u>If</u> $I_k(b_1) = z_1$; $I_k(b_2) = z_2$ $(z_1 \neq z_2)$,
b_1 and b_2 being two empirical manifestations and z_1 and z_2 two corresponding symbolic manifestations,

<u>and if</u> the empirical manifestations b_1 and b_2 are *undistinguishable* from the perspective of the representation target,

<u>then</u> I_k has a higher-than-necessary (or exaggerated) level of detail (resolution).

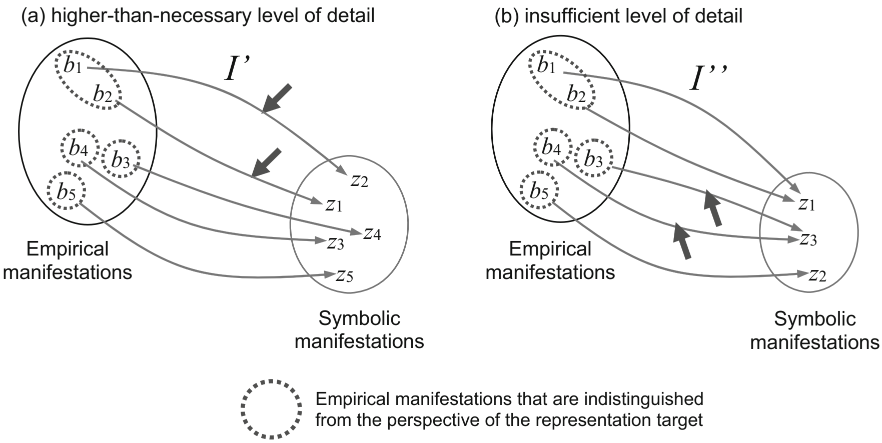

Fig. 4.7 Representation scheme of (**a**) an indicator with higher-than-necessary level of detail and (**b**) an indicator with insufficient level of detail

Example 4.16 A manufacturing process produces metal screws. The representation target is the "amount of production of the process". Indicator *I* represents the "weight of the screws produced daily". If the indicator's accuracy is ±1 g/day—when ±10 kg/day would be enough – then the level of detail will be higher than necessary.

In other words, if the indicator mapping is more accurate than required, two different empirical manifestations, which are indifferent according to representation target, can be unnecessarily distinguished (e.g., it would be meaningless to distinguish two daily productions (b_1 and b_2) with $I(b_1) = 653.321$ kg/day and $I(b_2) = 650.023$ kg/day).

Example 4.17 Let us consider the indicator represented in Fig. 4.8, which operationalizes the representation target "external design of a car". The scale "granularity" (i.e., 15 categories) is certainly excessive, considering the discriminatory power of ordinary respondents. In other words, this illusory high resolution does not necessarily imply accuracy in the respondents' evaluations, which are unlikely to be reliable and repeatable.

On the other hand, the resolution of one indicator could be lower than required (or insufficient), causing loss of important information on the process investigated. In general, the level of detail can be considered insufficient for those indicators that map different empirical manifestations into the same symbolic manifestation and these empirical manifestations should be distinguished according to the representation target (see Fig. 4.7b).

In formal terms:

<u>If</u> $I_k(b_1) = z_1$; $I_k(b_2) = z_2$, being $z_1 \approx z_2$
<u>and if</u> the empirical manifestations b_1 and b_2 are *distinguishable*, according to the representation target
<u>then</u> I_k has an insufficient level of detail

The insufficient level of detail (for single indicators) is closely related to the notion of (non-)exhaustiveness for sets of indicators, as discussed later in Sect. 4.6.2.

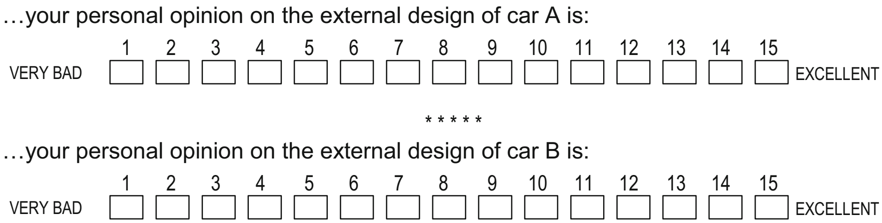

Fig. 4.8 Example of indicator with an exaggerated level of detail. The link between empirical manifestations (opinion on the external design of a car) and symbolic manifestations (scale categories) can be dubious and unreliable, due to the excessive number of categories

Counter-Productivity

In a process managed using indicators, managers and employees often focus their attention on indicators related to *local* incentives[2]; this behaviour can sometimes be counter-productive since it may negatively impact on the *global* performance of the whole process.

Example 4.18 The main purpose of a construction company is to reduce the indicator "construction time", in order to take a competitive advantage with respect to competitors. This behaviour may generate several counter-productive actions:

– to save time, employees do not obey safety rules (e.g., they do not use safety helmets/harness);
– working vehicles, rushing around the building site to save time, become dangerous for public safety;
– customer satisfaction decreases as the result of the work is poor, due to the excessive speed up.

Focusing too much on a single *dimension* of the process (i.e., "construction time", which could be represented by a dedicated indicator) can therefore be counter-productive in global terms.

Precisely, the notion of counter-productivity can be illustrated as follows. Some sub-indicators concur in representing different dimensions/features of a process. It is assumed that the *global* performance of the process can be monitored, perhaps through a derived indicator (I_{TOT}) that aggregates the sub-indicators. If the increase of a specific sub-indicator (I_k) is related to the decrease of (at least) one other sub-indicator (negative correlation), determining a significant decrease in the global performance (which can be depicted by the derived indicator I_{TOT}), then I_k is counter-productive. Of course, this definition entails that sub-indicators are defined on ordinal or more powerful scales, i.e., scales for which order relations are meaningful.

If counter-productive sub-indicators are linked to incentives, the attention of users may dangerously focus on them, to the detriment of global performance.

The notion of counter-productivity may also be defined in formal terms. Let suppose that a process is represented by a number of sub-indicators that are aggregated into a derived indicator I_{TOT}. Assuming that the process skips from condition "b_1" to condition "b_2", the increase of a sub-indicator (I_k) is correlated to the decrease of one or more sub-indicators (e.g., I_h, I_i, and I_l)[3]:

[2]E.g., internal benefits for managers and employees who contribute to increase some specific indicators.

[3]For simplicity, it has implicitly been supposed that an increase in the performance of a sub-indicator determines an increase in the corresponding value, and *vice versa*.

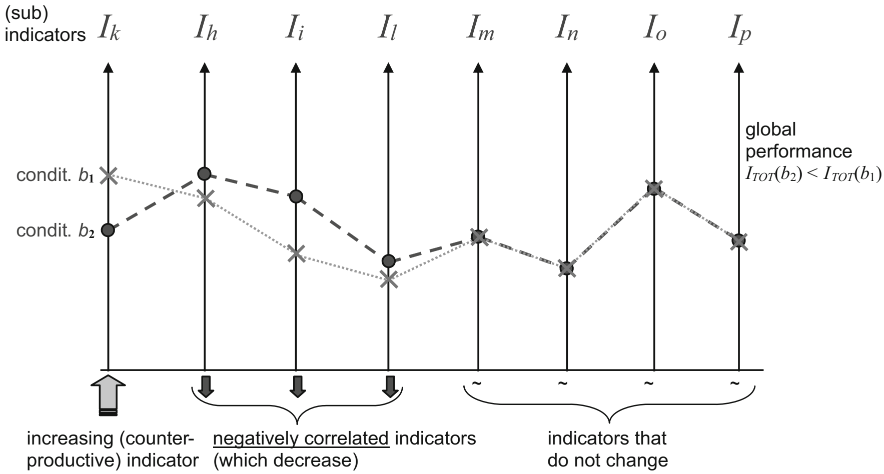

Fig. 4.9 Concept of counter-productive indicator. When passing from condition b_1 to b_2, the increase in one sub-indicator (I_k) produces a decrease in three other (negatively correlated) sub-indicators (I_h, I_i and I_l), with a consequent decrease in the global performance (depicted by I_{TOT})

$$\begin{gathered} I_k(b_2) > I_k(b_1) \\ I_h(b_2) < I_h(b_1) \\ I_i(b_2) < I_i(b_1) \\ I_l(b_2) < I_l(b_1) \\ \ldots \end{gathered}$$

If the above variations cause a reduction in the global performance, which is depicted by the derived indicator (i.e., $I_{TOT}(b_2) < I_{TOT}(b_1)$), then I_k can be classified as counter-productive. Figure 4.9 schematizes the concept of counter-productivity.

When testing the counter-productivity property, a crucial point is identifying (real or potential) correlations among sub-indicators.

Example 4.19 A call center uses several indicators to estimate customer satisfaction. Two of them are:

I_1 "average response time" (i.e., waiting time before a call is answered);
I_2 "percentage of unanswered calls".

These two indicators can be counter-productive because employees can "game" the process by answering a call immediately and then putting it on hold before starting the conversation. Although this behaviour makes I_1 and I_2 increase, it can be counter-productive from the perspective of other indicators of customer satisfaction, e.g., "number of exhaustive answers", "courtesy", "number of queued calls", etc.

Since the increase in I_1 and I_2 could negatively impact on the global performance of the call center, these indicators can be classified as counter-productive.

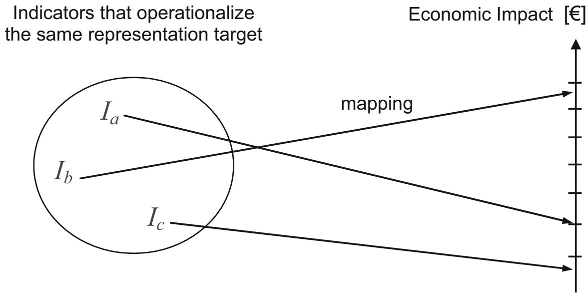

Fig. 4.10 Mapping performed to estimate the economic impact of a set of indicators

Economic Impact

The economic impact of an indicator strictly depends on the nature of the process investigated and can be studied in relative terms, e.g., comparing two (or more) alternative indicators that operationalize the same representation target. In general, we can hardly assert whether one indicator is economic (or not) in absolute terms, but we easily assert whether it is more (or less) economic than another one.

To compare different indicators, we have to set up a suitable mapping of their economic impact (see Fig. 4.10); this mapping is not necessarily unique but it depends on the nature of the process investigated. For instance, it could be based on the data-collection and data-processing cost.

Example 4.20 A small company produces punched metal components. Two possible indicators can be used to check the dimensional conformity of a cylindrical element:

I_1 "diameter measurement" obtained using an accurate calliper. Time needed is about 9 s.

I_2 "result of a (go—no go) manual testing", using a *hard gauge* with a calibrated shaft of the minimum tolerable diameter. Time needed is about 3 s.

Supposing that measurement cost is directly proportional to the time consumed, I_2 can be considered three times more economical than I_1.

Simplicity of Use

Similarly to the previous property, simplicity of use can be studied in relative terms, comparing two (or more) alternative indicators that operationalize the same representation target.

This comparison may include several complementary criteria concerned with *simplicity of use* (e.g., indicators should be easy to understand, easy to use, etc.).

Example 4.21 The simplicity of use of the two indicators (i.e., I_1 and I_2) defined in Example 4.20 can be evaluated considering the following criteria:

(a) technical skill required by operators;
(b) time required.

Fig. 4.11 Technical skill required by operators is evaluated using the scale: 1-low, 2-medium, 3-high difficulty; time required is evaluated using the scale: 1-short, 2-medium, 3-long time. Possible scheme to determine the simplicity of use of I_1 and I_2

	(a) technical skill (1 to 3)	(b) time required (1 to 3)	simplicity of use (sum)
I_1	2	3	2 + 3 = 5
I_2	1	1	1 + 1 = 2

These criteria are evaluated using two corresponding three-level ordinal scales (*low/medium/high difficulty* and *short/medium/long time*).

A very rough indication of the *simplicity of use* can be determined summing the scale level values: the smaller the sum, the simpler the indicator. Figure 4.11 shows that I_2 is simpler than I_1 as it requires a lower technical skill and less time.

It is interesting to note that the two criteria (a) and (b) are evaluated on a three-level ordinal scale, while *simplicity of use*—being obtained through a sum—is supposed to be defined on an interval scale (Stevens 1951). We are aware that this sort of "promotion" may sometimes produce paradoxical results; however, it can be tolerated when the purpose is just to provide a rough aggregation.

4.6.2 Properties of Sets of Indicators

A set of indicators is supposed to represent a process or a relevant portion of it. Indicators should represent the dimensions of the process, without omissions or redundancies. To this purpose, *exhaustiveness* and *non-redundancy* are two desirable properties. Before formalizing these properties, we provide some general definitions that will be used later in this manuscript.

Definition 4.1 A generic process can be in different states or conditions. The *state of a process* is a "snapshot" of the indicators in use, in a specific moment/situation.

Example 4.22 Three indicators represent the sales of a company:
I_1 "number of products sold daily" [units];
I_2 "daily takings" [€];
I_3 "daily profit" [€].

Two possible states are:
i-th day: $I_1(i) = 203$ units; $I_2(i) = 4820$ €; $I_3(i) = 3600$ €
j-th day: $I_1(j) = 178$ units; $I_2(j) = 5680$ €; $I_3(j) = 3546$ €

In this case, each state is a "snapshot" of the process in a particular day.

Complex processes are generally represented from the perspective of several *dimensions*, through several indicators. Selecting the relevant *dimensions* is a

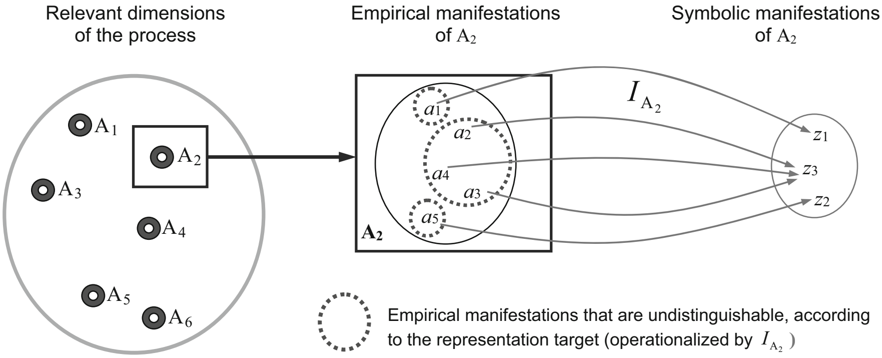

Fig. 4.12 Schematic representation of the concept of *set of indicators*. For each process dimension (A_1, A_2, A_3, ...), it is possible to define one or more indicators. All the indicators form a *set* or *family*

difficult task, because it requires a complete and thorough knowledge of the process of interest.

Definition 4.2 A *set* or *family* of indicators is a group of indicators that are supposed to represent a generic process (or part of it). For complex processes, indicators are generally grouped into sub-sets, depending on their affinity (see Fig. 4.12).

Exhaustiveness

Exhaustiveness is probably the most important property for a *set* of indicators that are supposed to represent a process.

A set of indicators is considered *non-exhaustive* when:

- The representation does not consider one or more (important) dimensions of the process; i.e., the set is *incomplete* since some indicators are missing (see Fig. 4.13).
- One or more indicators do not map distinguished empirical manifestations into distinguished symbolic manifestations; it can therefore be asserted that the mapping resolution is lower than necessary (see Fig. 4.14).

The property of exhaustiveness can also be explained as follows. If a set of indicators is unable to discriminate two process states ("*a*" and "*b*") with some distinguished empirical manifestations, then it is not exhaustive. This condition may be expressed in formal terms as follows:

<u>If</u> $\forall j \in F,\ I_j(a) \approx I_j(b)$

<u>and if</u> some empirical manifestations in state "*a*" are distinguished from those in state "*b*",

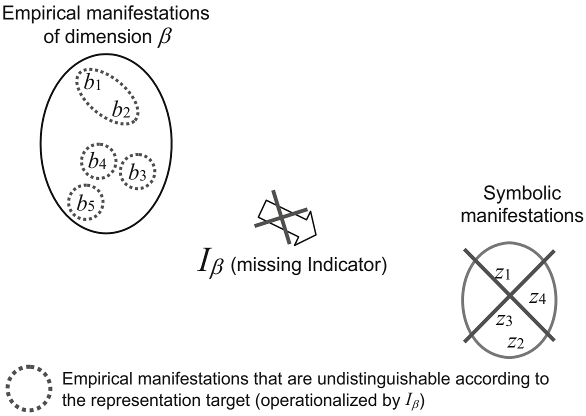

Fig. 4.13 Missing indicator: the representation does not include a process dimension and a corresponding indicator

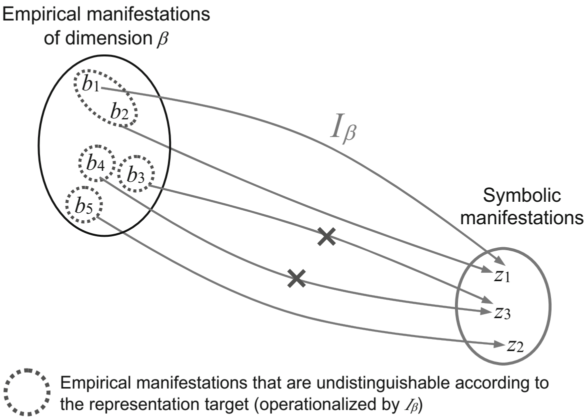

Fig. 4.14 Example of indicator with lower-than-necessary resolution: distinguished empirical manifestations are not distinguished by the indicator mapping

<u>then</u> the set of indicators is not exhaustive,

being:

a and *b* states of the process;
F set (family) of indicators.

Example 4.23 In a process manufacturing metal components, the following set of indicators is adopted:

I_1 "total number of units produced in one year" [units];
I_2 "manufacturing time" [h];
I_3 "lead time" (i.e. supply time, tool-change time, etc.) [h].

Fig. 4.15 Example of non-exhaustive set of indicators (I_1, I_2, I_3), which becomes exhaustive when adding a new indicator (I_4)

	I_1	I_2	I_3	I_4
State (*a*)	300,000 units	2700 h	700 h	2.1%
State (*b*)	300,000 units	2700 h	700 h	3.5%

non-exhaustive set of indicators

exhaustive set of indicators

These indicators are not able to distinguish two possible process states ("*a*" and "*b*"), which can be distinguished when introducing a new indicator, i.e., I_4, "percentage of defective units produced annually" [%] (see Fig. 4.15). The initial set is therefore non-exhaustive, while the new one (including I_4) is exhaustive.

Further Considerations on the Exhaustiveness Property

Exhaustiveness is a basic property for a good representation of a complex process, when using indicators (Roy and Bouyssou 1993).

Testing this condition requires a thorough knowledge of the process examined. Although the scientific literature does not include automatic tools to do it, it includes some popular approaches for the construction of a "balanced" set of indicators, such as the Balanced Scorecard or the EFQM (European Foundation for Quality Management) model, which will be described in Chap. 5.

Since a process is a dynamic system that is constantly evolving, representation targets may change over time. Indicators need to be periodically corrected and updated, in order to be aligned with representation targets, ensuring exhaustiveness of the representation (Flapper et al. 1996).

Non-redundancy

If (i) a set (or family) of indicators (F) is exhaustive and (ii) it continues to be exhaustive even when removing one indicator (I_k), then the latter indicator is *redundant*.

In formal terms:

<u>If</u> F fulfils the property of exhaustiveness
<u>and if</u> $\exists\, I_k \in F$: $F \backslash \{I_k\}$ keeps fulfilling exhaustiveness
<u>then</u> I_k is a redundant indicator

being:

F initial set of indicators;
$F \backslash \{I_k\}$ set of indicators, excluding I_k.

Fig. 4.16 Scheme of the concept of "redundant indicator"

if	I_1	I_2	I_3	I_4	I_5	is an exhaustive set of indicators
and if	I_1	I_2		I_4	I_5	is a (residual) set that continues to be exhaustive
then			I_3			is a redundant indicator

Example 4.24 In a manufacturing company producing plastic component, the process is represented through a set of four indicators:

I_1 "total number of units produced annually";
I_2 "number of defective units produced annually";
I_3 "manufacturing time";
I_4 "production efficiency", calculated as: $I_4 = \frac{I_3 - I_5}{I_3}$;
I_5 "down times", such as supply time, tool change time, repairing time, etc.

It is assumed that the initial set of indicators fulfils the property of exhaustiveness; then, the indicator I_3 is removed from the set. If the (residual) set (I_1, I_2, I_4, I_5) continues to be exhaustive, then I_3 is classified as redundant (see Fig. 4.16).

Usually, indicators that can be deduced from other ones − e.g., I_3, which is a function of I_4 and I_5 − are likely to be redundant. The presence of redundant indicators does not significantly contribute to enriching the process representation.

From the perspective of the *representation theory of indicator*, an indicator is classified as redundant when the empirical manifestations that it maps are already considered by other indicators, or if it is useless for representation.

4.6.3 Properties of Derived Indicators

A *derived* indicator transforms and/or aggregates the information of the sub-indicators into a single synthetic indicator. For this reason, derived indicators may contribute to simplify the process representation.

Example 4.25 Likewise Example 4.11, the pollution level of the exhaust of a motor vehicle is estimated considering four *basic* indicators, representing the concentrations of four relevant pollutants (I_x). The concentration of each pollutant is then mapped into a five-level sub-indicator (I'_x). Let us consider two possible states (a and b), with the aim of determining the worst one (Fig. 4.17).

To evaluate global pollution, it is convenient to define a *derived* indicator (I), which aggregates the information of the previous ones. I can be defined as the maximum of the four (sub-)indicators values:

	I'_{NO_x}	I'_{SO_2}	I'_{CO}	$I'_{PM_{10}}$
State (a)	4	4	3	4
State (b)	1	1	1	5

Fig. 4.17 Comparison between two different pollution levels of the exhaust emissions of a motor vehicle

$$I(a) = \max\left\{I'_{NO_X}, I'_{HC}, I'_{CO}, I'_{PM_{10}}\right\} = \max\{4, 4, 3, 4\} = 4 \tag{4.15}$$

$$I(b) = \max\left\{I'_{NO_X}, I'_{HC}, I'_{CO}, I'_{PM_{10}}\right\} = \max\{1, 1, 1, 5\} = 5 \tag{4.16}$$

According to I, state (b) is worse than state (a).

The aggregation of indicators can considerably simplify the analysis but can also be questionable or misleading (Franceschini et al. 2006a). For instance, state (b) is considered more dangerous than state (a), even if the risk level of three pollutants (I'_{NO_2}, I'_{SO_2}, and I'_{CO}) is much lower. Is this correct, from the point of view of human health?

The rest of this section will illustrate some properties that may support the aggregation of sub-indicators into derived indicators.

Property of (Strict/Weak) Monotony

This property concerns a *derived* indicator that aggregates a set of sub-indicators. Simplifying, if the increase/decrease of a specific sub-indicator is not associated to the increase/decrease of the derived indicator (*ceteris paribus*), then the derived indicator is not (strictly) *monotonous* with respect to that specific sub-indicator.

This definition implicitly entails that the symbolic manifestations of the sub-indicators are represented using ordinal or more powerful scales. When indicators are represented on scales with no order relation (i.e., nominal scales), the property of monotony loses its meaning.

We remark that monotony should always be expressed with respect to a specific sub-indicator: while it makes sense to state that a derived indicator is (or not) monotonous with respect to a certain sub-indicator, it makes no sense to state that a derived indicator is monotonous in general.

Going into more detail, the property of monotony is closely linked to the definition of monotonous function in Mathematical Analysis (Hazewinkel 2013). Precisely, a derived indicator $f(x)$ is called *strictly monotonically increasing* with respect to a (sub)indicator, if—for all x and y values of this (sub)indicator such that $x < y$—one has $f(x) < f(y)$ (i.e., f preserves the order). Likewise, a derived indicator is called *strictly monotonically decreasing*, if—for all x and y such that $x < y$—one has $f(x) > f(y)$ (i.e., f reverses the order).

If the strict order "$<$" in the previous definition is replaced by the weak order "$\leq$", then one obtains a weaker requirement. A derived indicator that meets this requirement is called *weakly monotonically increasing*. Again, by inverting the

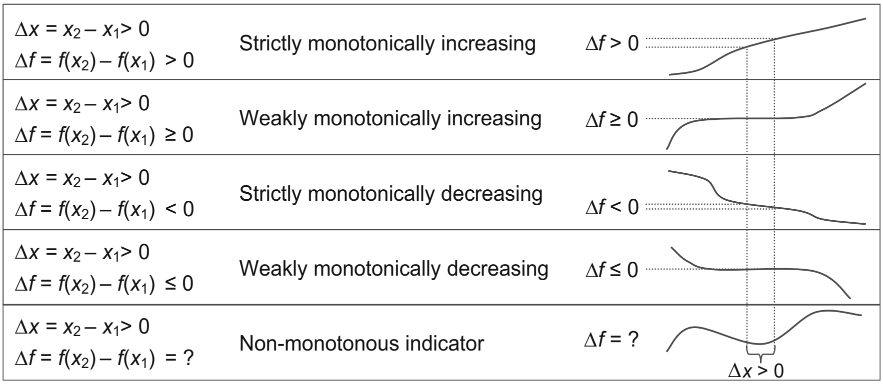

Fig. 4.18 Schematic representation of the property of (strict/weak and increasing/decreasing) monotony of a derived indicator (f) with respect to a sub-indicator (x). For simplicity, $f(x)$ is represented as a continuous function

order symbol, one finds a corresponding concept called *weakly monotonically decreasing* (see Fig. 4.18).

Derived indicators that are strictly monotonous with respect to a (sub)indicator are *one-to-one* (because for $x \neq y$, either $x < y$ or $x > y$ and so, by monotony, either $f(x) < f(y)$ or $f(x) > f(y)$, thus $f(x) \neq f(y)$.

In general, (strict) monotony is a desirable property of derived indicators, since it proves their responsiveness with respect to variations in the relevant sub-indicators. Strict monotony is therefore preferred to weak monotony, which is in turn preferred to non-monotony.

Example 4.26 Returning to Example 4.15 about FMECA (*Failure Mode, Effects and Criticality Analysis*), it can be demonstrated that the derived indicator RPN_i is strictly monotonically increasing with respect to each of the three sub-indicators S_i, O_i and D_i. A simple mathematical demonstration follows

- Consider a generic failure mode for which $RPN_i = S_i \cdot O_i \cdot D_i$;
- If S_i increases ($\Delta S_i > 0$), the new value of the derived indicator will be $RPN'_i = S'_i \cdot O_i \cdot D_i = (S_i + \Delta S_i) \cdot O_i \cdot D_i = RPN_i + \Delta S_i \cdot O_i \cdot D_i$;
- Since O_i and D_i are both positive (as they are natural numbers $\in [1, 10]$), the term $\Delta S_i \cdot O_i \cdot D_i > 0$ and therefore $RPN'_i > RPN_i$.

Similar considerations can be adapted to prove the strict monotony of RPN_i with respect to O_i and D_i.

Example 4.27 The pollution level of the exhaust of a motor vehicle is estimated using the following aggregation model (see also Example 4.11):

$$I_{TOT}^{A} = \max\left\{I'_{NO_X}, I'_{HC}, I'_{CO}, I'_{PM_{10}}\right\} \tag{4.17}$$

Table 4.6 Example of a weakly monotonically increasing *derived* indicator (I_{TOT}^{A}). Although the sub-indicator I'_{NO_X} shifts from 1 to 3, I_{TOT}^{A} does not change

	I'_{NO_X}	I'_{HC}	I'_{CO}	$I'_{PM_{10}}$	I_{TOT}^{A}
State (*a*)	1	1	1	3	3
State (*b*)	3	1	1	3	3

Assuming that one of the sub-indicators (i.e., I'_{NO_X}) increases when passing from state *a* to state *b*, the derived indicator (I_{TOT}^{A}) will not necessarily increase (see Table 4.6). In fact, it can be easily demonstrated that I_{TOT}^{A} is a weakly monotonically increasing (derived) indicator with respect to each of the sub-indicators. In other terms, I_{TOT}^{A} does not necessarily "respond" to any variation in sub-indicators.

The example shows that using a derived indicator which is not strictly monotonous, we may lose some information (according to I_{TOT}^{A}, there is no difference between state *a* and state *b*). In Sect. 2.2, we have described other indicators that are not strictly monotonous (e.g., *ATMO*, *AQI* and *IQA*).

Example 4.28 Oversimplifying, Bibliometrics is a branch of science studying and developing quantitative methods to measure the performance of a generic research system. i.e., a scientist, a research institution, a group of researchers, a scientific journal, etc. (Franceschini and Maisano 2017).

The so-called Hirsch index (*h*) has been proposed to measure both the *productivity* and citation *impact* of the scientific output of a scientist. This indicator is defined as "the number such that, for a general group of scientific papers (e.g., those of a scientist), *h* papers obtained at least *h* citations while the other papers received no more than *h* citations" (Hirsch 2005; Franceschini and Maisano 2010). In the example in Fig. 4.19, $h = 7$ since seven publications obtained at least seven citations each.

We now analyse the Hirsch-index, demonstrating the following statements.

(i) This indicator cannot decrease over time, i.e., it is weakly monotonically increasing with respect to time.
(ii) The *h*-index of a scientist is limited by: (1) the number of citations of the more cited papers, and (2) the total number of papers.

Paraphrasing the statement at point (i), the *h* index is (supposed to be) weakly monotonous with respect to time. This condition is verified because the number of scientific articles and corresponding citations—which both contribute to raising *h*—cannot decrease over time. For example, in the event of career interruption or retirement of a scientist, his/her *h* will remain constant or even increase (e.g., some of the already published papers obtain new citations).

The statement at point (ii) can be deduced from the so-called Ferrers diagrams, which provide a visual representation of the papers and the relevant citations (Franceschini and Maisano 2010). For example, the Ferrers diagram in Fig. 4.20

	Citations of each paper	Paper rank (decreasing no. of citations)
h subset	30	1
	20	2
	18	3
	12	4
	9	5
	8	6
	8	7
	6	8
	6	9
	5	10
	…	…

Fig. 4.19 Example of calculation of the *h*-index

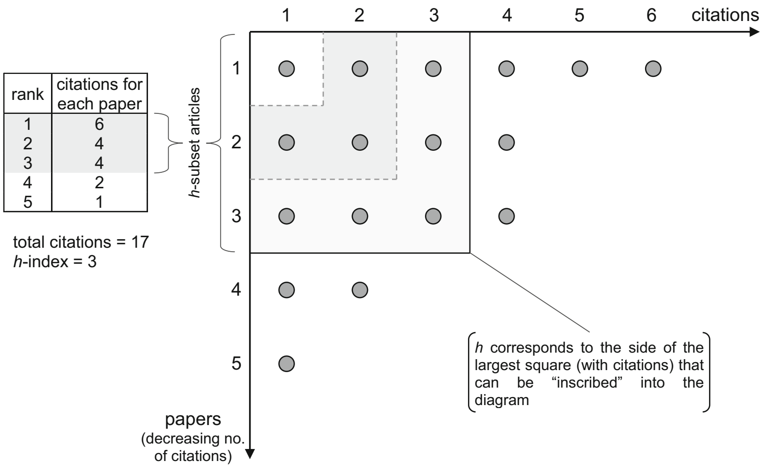

Fig. 4.20 Example of a Ferrers diagram related to the papers/citations of a fictitious scientist. Adapted from Franceschini and Maisano (2010). With permission

shows that the growth of *h* (corresponding to the side of top-left square) is limited by (1) the number of papers (in the vertical axis), and (2) the number of citations of the more cited papers (in the horizontal axis).

Property of Compensation

The property of compensation can be studied when a process (or a portion of it) is represented by sub-indicators that are aggregated into a *derived* indicator. In a nutshell, if variations in two or more sub-indicators may compensate each other—without any variation in the *derived* indicator—then the derived indicator fulfils the property of compensation.

The property can be formalized as follows. Let us consider two sub-indicators (I_1 and I_2), which are aggregated into a derived indicator (D).

If (first condition) a variation in I_1 (i.e., ΔI_1) always causes a variation in D (i.e., ΔD), *ceteris paribus*,

and if (second condition) there exist a variation in I_2 (i.e., ΔI_2) that compensates for the previous ΔD (i.e., $\Delta D = 0$),

then D fulfills the property of compensation and a *substitution rate* between I_1 and I_2 can be univocally determined.

Compensation is a typical property of *additive* and *multiplicative* aggregation models.

Returning to the formal definition, we note that the first condition is automatically satisfied if D is strictly monotonic with respect to I_1 (sufficient but not necessary condition).

As for the *substitution rate*, it is defined as the portion (ΔI_1) of the first sub-indicator (I_1) that one should give up in order to obtain a variation (ΔI_2) in the second sub-indicator (I_2), keeping the derived indicator (D) constant. The substitution rate can be also seen as an analytic function connecting ΔI_1 to ΔI_2. Depending on the type of aggregation of sub-indicators, the substitution rate may be constant or it may depend on the so-called *operating point*, i.e., the initial values of sub-indicators I_1 and I_2. Examples 4.29 and 4.30 show how to calculate the substitution rate in an exact or approximate way. These examples also show that constant substitution rates are generally preferable as they make the mutual compensation of sub-indicators more controllable.

Further comments on the property of compensation follow:

- For simplicity, the previous definition of compensation is referred to two sub-indicators only. With appropriate adjustments, the same property can be extended to three or more sub-indicators.
- There may be derived indicators with more *relaxed* forms of compensation, than the one formalized previously. In general, we can define as *weak* compensation a form of compensation in which variations in two (or more) sub-indicators may compensate each other, not necessarily producing any variation in the derived indicator.
- The fact that an indicator follows the property of (weak) compensation or it does not follow it at all may depend on (i) the aggregation model of the derived indicator or (ii) any constraints in the sub-indicator scales. These concepts are clarified by the following four examples.

Example 4.29 A company develops a (derived) indicator to evaluate the global performance of a production line, denominated "Overall Equipment Efficiency". This indicator should aggregate three sub-indicators, which are supposed to depict three dimensions:

- *System availability*: A = operating time / scheduled time;
- *Performance productivity*: B = (units produced · ideal unitary cycle time) / operating time;
- *Quality*: C = (Units produced − defective units) / Units produced.

It can be noticed that the variability range of each of the three sub-indicators is included between 0 and 1.

There are two options for aggregating these sub-indicators:

$$\begin{aligned}&(1)\ OEE_1 = A \cdot B \cdot C\\ &(2)\ OEE_2 = \frac{A+B+C}{3},\end{aligned} \tag{4.18}$$

i.e., the former relies on a multiplicative model while the latter on an additive one.

We now analyse OEE_1 and OEE_2, answering the following questions:

(i) Do these indicators fulfil the compensation property.
(ii) If so, what are the relevant substitution rates between A and B?
(iii) Do these substitution rates depend on the so-called "operating point"?
(iv) Determine the ΔB values that compensate for a variation of $\Delta A = -1\%$, when $A = 70\%$, $B = 30\%$ and $C = 80\%$.

Both the derived indicators (OEE_1 and OEE_2) fulfil the property of compensation with respect to each of the sub-indicators, since the following two conditions are met:

1. The variation of any one of the three sub-indicators (*ceteris paribus*) always produces a variation of the derived indicator;
2. The variation of one of the three sub-indicators can be compensated by that of (at least) another sub-indicator, so that the derived indicator has no variation.

Regarding OEE_1, the *substitution rate* between A and B can be determined by expressing B as a function of OEE_1, A and C:

$$B = \frac{OEE_1}{A \cdot C}. \tag{4.19}$$

Differentiating B with respect to A, we obtain:

$$dB = \frac{OEE_1}{A^2 \cdot C} dA = -\frac{A \cdot B \cdot C}{A^2 \cdot C} dA = -B \cdot \frac{dA}{A}. \tag{4.20}$$

We remark that the above expression can be used in the case of infinitesimal variations of the sub-indicators A and B only.

In the more general case of finite variations, the substitution rate between A and B can be determined by imposing that, in the presence of variations of sub-indicators (ΔA and ΔB), the variation of OEE_1 is null. Translating into formulae:

$$\Delta OEE_1 = OEE_1{}' - OEE_1 = (A + \Delta A) \cdot (B + \Delta B) \cdot C - A \cdot B \cdot C = 0, \quad (4.21)$$

from which:

$$\Delta B \cdot (A + \Delta A) = -B \cdot \Delta A \qquad \Rightarrow \qquad \Delta B = \frac{-B \cdot \Delta A}{A + \Delta A}. \quad (4.22)$$

It can be noticed that this substitution rate and also the one in Eq. (4.20) depend on the so-called "operating point", i.e., the initial values of sub-indicators A and B.

Regarding OEE_2, the substitution rate between A and B can be determined by expressing B as a function of OEE_1, A and C:

$$B = 3 \cdot OEE_2 - A - C. \quad (4.23)$$

Differentiating B with respect to A, we obtain:

$$dB = -dA. \quad (4.24)$$

The above expression is determined assuming that the variations of sub-indicators A and B are infinitesimal.

In the more general case of finite variations, the substitution rate between A and B can be determined by imposing that, in the presence of variations of sub-indicators (ΔA and ΔB), the variation of OEE_2 is null. Translating into formulae:

$$\Delta OEE_2 = OEE_2{}' - OEE_2 = \frac{(A + \Delta A) + (B + \Delta B) + C}{3} - \frac{A + B + C}{3} = 0,, \quad (4.25)$$

from which:

$$\Delta B = -\Delta A. \quad (4.26)$$

It can be noticed that this substitution rate perfectly follows that in Eq. (4.24) and does not depend on the "operating point".

Considering a specific situation in which $A = 70\%$, $B = 30\%$ and $C = 80\%$, the corresponding OEE_1 and OEE_2 values will be respectively:

$$\begin{aligned} OEE_1 &= A \cdot B \cdot C = 70\% \cdot 30\% \cdot 80\% = 16.8\% \\ OEE_2 &= \frac{A + B + C}{3} = \frac{70\% + 30\% + 80\%}{3} = 60\% \end{aligned}. \quad (4.27)$$

It can be noticed that OEE_2 is about one order of magnitude higher than OEE_1.

Assuming that $\Delta A = -1\%$, the relevant ΔB variation that compensate for it can be obtained by applying Eqs. (4.22) and (4.26) respectively:

$$\begin{array}{l}(OEE_1) \Rightarrow \Delta B = \dfrac{-B \cdot \Delta A}{A + \Delta A} = \dfrac{-30\% \cdot (-1\%)}{70\% + 1\%} = 0.4225\% \\ (OEE_2) \Rightarrow \Delta B = -\Delta A = 1\% \end{array}. \quad (4.28)$$

This result shows that the use of different aggregation models can produce very different results. For a careful examination of the *pros* and *contras* of additive/multiplicative models, we refer the reader to Joint Research Centre-European Commission (2008).

Example 4.30 Returning to Example 4.15 about FMECA (*Failure Mode, Effects and Criticality Analysis*), the derived indicator $RPN_i = S_i{\cdot}O_i{\cdot}D_i$ fulfils the property of compensation with respect to each of the sub-indicators, since the following conditions are both met:

1. A variation of any one of the three sub-indicators (*ceteris paribus*) always produces a variation of RPN_i;
2. The variation of one of the three sub-indicators can be compensated by that of (at least) another sub-indicator, so that RPN_i has no variation.

The *substitution rate* between two sub-indicators (e.g., S_i and O_i) is determined by imposing that, in the case they have variations (ΔS_i and ΔO_i), the variation of RPN_i is null. Translating into formulae:

$$\Delta RPN_i = RPN_i' - RPN_i = (S_i + \Delta S_i) \cdot (O_i + \Delta O_i) \cdot D_i - S_i \cdot O_i \cdot D_i = 0, \quad (4.29)$$

from which:

$$\Delta O_i \cdot (S_i + \Delta S_i) = -O_i \cdot \Delta S_i \qquad \Rightarrow \qquad \Delta O_i = \frac{-O_i \cdot \Delta S_i}{S_i + \Delta S_i}. \quad (4.30)$$

This substitution rate depends on the so-called "operating point", i.e., the initial values of sub-indicators S_i and O_i.

To be precise, compensation would be limited by the actual domain of sub-indicators—i.e., consisting of natural numbers between 1 and 10. For example, assuming that for a certain failure mode $S_i = O_i = D_i = 1$ and $\Delta S_i = +1$, a possible compensation would require $\Delta O_i = -0.5$ and therefore $O'_i = 0.5$, which it is not possible, since O'_i cannot be a decimal number.

Example 4.31 In the field of bibliometric evaluations, *scientific output* is typically given by scientific publications, such as research papers, monographs, conference proceedings, etc. (see also Example 4.28). The two major features associated to scientific output of scientists are:

- number of publications (1, 2, 3, . . ., P), which is an indicator of *productivity*;
- number of citations related to each publication (c_1, c_2, c_3, . . ., c_P), which is an indicator of *diffusion*.

The propensity to cite empirically changes from field to field (e.g., publications tend to accumulate many more citations in Biology/Medicine than in Engineering/Mathematics).

In Bibliometrics, many (derived) indicators have been proposed to synthesize the two previous dimensions (publications and citations) into a single number. One of these indicators is the (so-called) *success*-index (Franceschini et al. 2012).

Considering a group of publications (typically those relating to a scientist), a single article is classified as "successful" if the number of citations obtained is greater than or equal to a specific threshold (T), representing the average citation potential for a publication within the discipline of interest—i.e., the number of citations that an article is likely to obtain in that field (see the example in Table 4.7).

We now analyse the *success*-index, answering the following questions.

(i) On which type of measurement scale is it defined?
(ii) Is it meaningful to say that "the effort to raise the *success*-index from 2 to 5 is equivalent to that of raising it from 45 to 48"?
(iii) Is this (derived) indicator monotonous with respect to the relevant sub-indicators? If so, strictly or weakly monotonous?
(iv) Does this indicator fulfil the property of compensation? If so, what is the relevant substitution rate?
(v) Is the indicator suitable to evaluate the performance of multi-disciplinary institutions (e.g. a university with scientists from multiple disciplines).
(vi) Is the indicator suitable for comparing scientists of different age?
(vii) What are the main drawbacks of the aggregation performed by this indicator?

The *success*-index is defined on a *ratio* scale, whose zero is not arbitrary and corresponds to the absence of the manifestation of the measured characteristic (i.e., the number of "successful" publications).

Since the statement of interest (i.e., "the effort to raise the *success*-index from 2 to 5 is equivalent to that of raising it from 45 to 48") concerns the distances between objects (i.e., *success*-indices) that are represented on a ratio scale, it will be

Table 4.7 Example of calculation of the *success*-index for a group of publications

Publication no.	c_i	T	Successful
1st	3	3.7	✗
2nd	4	3.7	✓
3rd	15	3.7	✓
4th	0	3.7	✗
5th	1	3.7	✗
$P = 5$	$C = \Sigma c_i = 23$		*Success*-index = 2

P is the total number of publications of interest
c_i is the number of citations received by the i-th publication
C is the total number of citations received by the publications of interest
T is a specific threshold for determining a successful publication

meaningful. We also note that the statement would be meaningful even if *success*-index were defined on an interval scale.

Regarding the monotony property, we clarify that sub-indicators are given by: (1) the number scientific publications and (2) the number of citations for each publication. However, the increase of one of these sub-indicators is not necessarily associated with the increase of the *success*-index (e.g., consider a new publication without citations or a new citation to an already successful publication). As a consequence, the *success*-index is a weakly monotonous indicator with respect to the aforementioned sub-indicators.

The *success*-index does not fulfil the compensation property since the variation of a sub-indicator is not necessarily associated with a change in the indicator itself. For this reason, the substitution rate makes no sense in this case and cannot be determined.

The *success*-index is potentially suitable to evaluate the performance of multi-disciplinary institutions, provided that the citations of publications relating to different scientific disciplines are compared with dedicated thresholds (T), which should reflect the relevant propensity to receive citations by the different disciplines (see the example in Table 4.8).

Additionally, for a fair comparison among institutions (e.g., universities) of different *size* their *success*-indices could be normalized dividing them by the corresponding number of scientists (Franceschini et al. 2012).

Regarding the comparability of scientists of different age, the *success*-index would be in favour of the ones with longer careers, because of the higher time available to produce publications and accumulate citations. Assuming that the scientific production of scientists is supposed to be proportional to their working

Table 4.8 Example of calculation of the *success*-index for a group of publications from three different scientific disciplines

Discipline	Publication no.	c_i	T	Successful
Biology	1st	15	13.3	✓
	2nd	21	13.3	✓
	3rd	12	13.3	✗
	4th	8	13.3	✗
Mathematics	5th	2	1.6	✓
	6th	0	1.6	✗
	7th	1	1.6	✗
Physics	8th	3	2.2	✓
	9th	1	2.2	✗
	$P = 9$	$C = \Sigma c_i = 63$		*Success*-index = 4

Note that the value of T can vary from discipline to discipline
P is the total number if publications of interest
c_i is the number of citations received by the i-th publication
C is the total number of citations received by the publications of interest
T is a specific threshold for determining a successful publication

Table 4.9 *Derived* indicator fulfilling the property of compensation

	I_{NO_X}	I_{HC}	I_{CO}	$I_{PM_{10}}$	I^B_{TOT}
State (*a*)	2	2	1	3	(2 + 2 + 1 + 3)/4 = 2
State (*b*)	1	1	3	3	(1 + 1 + 3 + 3)/4 = 2

age (e.g., in terms of career years), the *success*-indices of scientists could be normalized dividing them by the relevant ages.

The synthesis mechanism of the *success*-index has some drawbacks. The first one is represented by the fusion of sub-indicators (i.e., number of publications and relevant citations) into a single number, with the inevitable loss/degradation of information available. For example, the *success*-index does not take into account the so-called "excess" citations, i.e., the additional citations that an already-successful publication obtains. Furthermore, the determination of threshold values (T) can be rather delicate/sophisticated (Franceschini et al. 2012).

Example 4.32 With reference to Example 4.11, where the pollution level of motor vehicle exhaust emissions is estimated, we consider I^B_{TOT} as the synthesis indicator:

$$I^B_{TOT} = \frac{\left(I''_{NO_X} + I''_{HC} + I''_{CO} + I''_{PM_{10}}\right)}{4} \tag{4.31}$$

As illustrated in Table 4.9, the pollution level skips from state $S(a)$ to state $S(b)$. The decreases of I_{NO_X} and I_{HC} are compensated by the increase of I_{CO}. I^B_{TOT} value does not change.

Sections 2.2 and 2.3 show additional examples of compensation for other derived indicators (e.g., HDI, air quality indicators, etc.).

4.7 Accessory Properties

We have so far illustrated several properties to support the analysis of indicators. However, before defining indicators, representation targets that are consistent with the strategy of the organization of interest should be identified. According to Kaplan and Norton (1996) the link between strategies and indicators should consider four different perspectives (financial, customer, internal business process, learning and growth); each perspective should be linked to reasonable representation targets.

The following two accessory properties are introduced to help identifying representation targets that are consistent with the strategic objectives. These properties are classified as "accessory" since they are focused on representation targets, rather than indicators.

- *Long-term goals*. Since indicators should encourage the achievement of long-term goals of one organization (or part of it), representation targets should concern dimensions that are strictly related to these goals.
- *Customer orientation*. In a competitive market, customer satisfaction is supposed to be one of the main goals of organizations. Many indicators focus on internal needs such as throughput, staff efficiency, cost reduction, and cycle time. While these needs are all laudable, they usually have little direct impact on customers needs. So, it is important to identify process aspects with a strong impact on customer satisfaction. Quality Function Deployment (QFD) and a quality-design approaches can be valid tools to reach this objective (Franceschini 2001).

Example 4.33 Let us consider the problem of evaluating suppliers of a manufacturing company. Before defining a set of suitable indicators, it is necessary to identify the strategic representation targets. A first choice could be the representation target "identification of the suppliers with the more advantageous pricing terms", which can be operationalized by the "supply cost" indicator. Unfortunately, this choice is myopic as it is not in line with the aforesaid accessory properties. In fact, other representation targets and relevant indicators concerned with supplier evaluation may deserve attention, as they are potentially related with the achievement of *long-term goals* and/or *customer satisfaction*. E.g.:

1. "Replenishment lead time";
2. "On-time performance";
3. "Supply flexibility";
4. "Delivery frequency";
5. "Supply quality";
6. "Inbound transportation cost";
7. "Information coordination capability";
8. "Design collaboration capability";
9. "Exchange rates, taxes";
10. "Supplier viability".

4.8 Construction of Indicators and Property Checking

After having illustrated the major properties of performance indicators, we now present an operational method to construct and check indicators, consistently with these properties. Summarizing, the method is based on the following steps (see flow chart in Fig. 4.21):

- define the *process* and identify the characteristic *dimensions*;
- identify the *representation targets* related to each dimension;
- analyse the consistency of representation targets with the strategic objectives of the organization of interest (i.e., testing *accessory properties*);
- define indicators preliminarily;

- for each indicator, check the *consistency with representation target*;
- for sets of indicators, check the properties of *exhaustiveness* and *non-redundancy*;
- define the measurement scale and the data-collection procedure of each indicator; check the remaining *properties of single indicators*;
- check the properties of derived indicators (*monotony* and *compensation*).

This method is based on a "top-down" testing. After having identified the major process dimensions, representation target(s) are identified consistently with the strategies of the organization (*accessory properties*). Then, a preliminary definition of possible pertinent (sets of) indicators is performed. For each indicator, we should make sure that it well represents a particular representation target (property of *consistency with representation target*). Next steps include (i) testing the properties

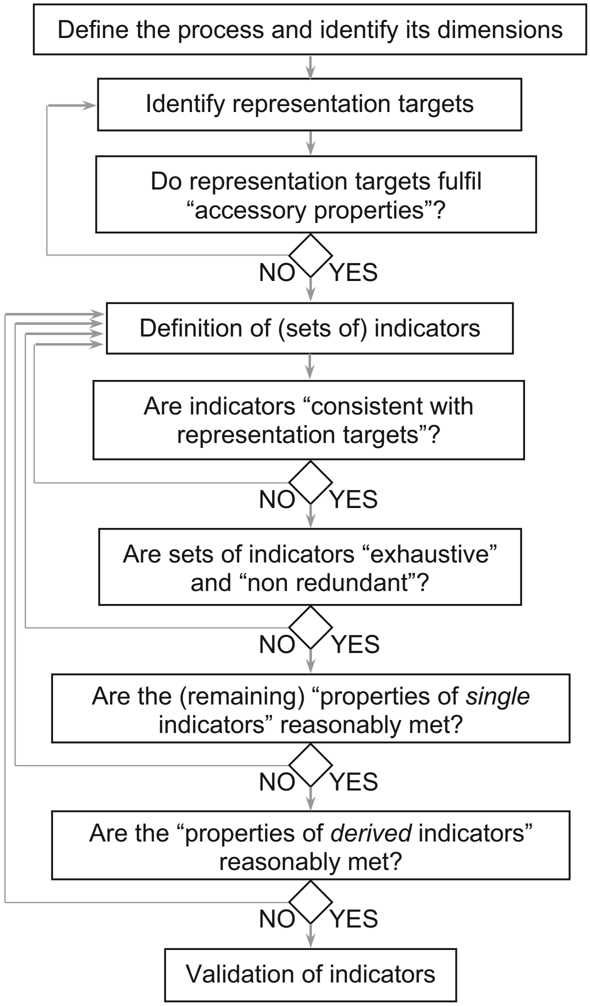

Fig. 4.21 Scheme of the suggested operational method

of the sets of indicators (*exhaustiveness* and *non-redundancy*) and (ii) testing the remaining properties of single indicators (*meaningfulness of statements*, *level of detail*, *counter-productivity*, *economic impact*, and *simplicity of use*).

After having tested the *properties of single indicators*, we should check the criteria for aggregating sub-indicators, the scale properties of derived indicators, and their properties of *monotony* and *compensation*.

As illustrated in Fig. 4.21, the proposed method may require several iterations, denoted by several feedback loops (e.g., definition, test, correction, redefinition, etc.), before developing a proper system of indicators. However, this is a very important task for identifying the potential drawbacks and misuses of indicators.

In the remainder of this chapter, we complete the previous description by means of a realistic application example. The next chapter will deal with the problem of designing a performance measurement system, in a more detailed and structured way.

Example 4.34 An *automotive* company should define a system of indicators to evaluate the overall performance of a car. The analysis is articulated into the following points:

- *Process Identification*

The major *dimensions* of the process of interest are:

1. Technical features;
2. Running-dynamics performance;
3. Fuel consumption;
4. Polluting emissions;
5. Comfort;
6. Material cost;
7. Production cost.

- *Identification of representation targets*

For each process dimension, representation targets should be identified (see Table 4.10).

- *Representation-target analysis and testing*

Representation targets should be analysed and tested according to the "accessory properties" (see Table 4.11). If necessary, they should be adjusted consistently with these properties.

Table 4.10 List of the major representation targets related to the process dimensions

Process dimension	Representation target(s)
1. Technical features	1.1—Car mass
	1.2—Engine characteristics
	1.3—Space available to passengers
2. Running-dynamics performance	2.1—Engine power
	2.2—Acceleration
	2.3—Cornering stability
	2.4—Braking performance
3. Fuel consumption	3.1—Average fuel consumption
4. Polluting emissions	4.1—Pollution level of exhaust emissions
5. Comfort	5.1—Car-seat comfort
	5.2—Vibrations
	5.3—Noise
6. Material cost	6.1—Cost of raw materials
	6.2—Purchase cost of automotive subsystems from third parties (e.g., seats, brakes, air-conditioner, etc.)
7. Production cost	7.1—Manufacturing cost
	7.2—Assembly cost
	7.3—Logistics cost

Table 4.11 Accessory properties for testing representation targets

a—Long-term goals	a_1—Car components must not deteriorate rapidly; a_2—The product should meet the current (and future) regulations for polluting emissions.
b—Customer orientation	b_1—Car performance should be competitive;
	b_2—Main focus on passengers' safety and comfort.

- *Preliminary definition of indicators*

The next step is a detailed definition of indicators (Table 4.12). These indicators include one *derived* indicator (4.1.5) and one *subjective* indicator (5.1.1). Most of the proposed indicators are measurements, except the latter ones and those related to cost items (i.e., 6.1.1–7.3.1).

Then the *consistency with representation target* of indicators is tested, trying to answer the following question: "Is the representation target properly operationalized by the proposed indicator?".

- *Testing the sets of indicators*

Having identified the sets of indicators, we should test their *exhaustiveness*. To this purpose, each process *dimension* is analysed separately, considering the relevant representation targets.

Table 4.12 Initial list of the proposed indicators

Representation target	Indicators
1.1—Car mass	1.1.1—Car mass in medium-load conditions.
1.2—Engine characteristics	1.2.1—Engine displacement [cm^3].
1.3—Space available to passengers	1.3.1—Interior volume [dm^3].
2.1—Engine power	2.1.1—Maximum engine power [kW] (bench test).
2.2—Acceleration	2.2.1—Average acceleration from 0 to 100 km/h.
2.3—Cornering stability	2.3.1—Maximum speed along a predefined curve (in safe conditions).
2.4—Braking performance	2.4.1—Braking space for stopping the car from the speed of 100 km/h in normal conditions (dry road, moderately worn tyres and medium car load); 2.4.2—Average deceleration, in the above conditions.
3.1—Average fuel consumption	3.1.1—Average consumption on three standard tracks (urban/suburban/mixed).
4.1—Pollution level of exhaust emissions	4.1.1 to 4.1.4—Concentration of four pollutants, expressed using the indicators in Example 4.11; 4.1.5—Derived indicator summarizing the global pollution level (see Example 4.11).
5.1—Car-seat comfort	5.1.1—Indicator of customer satisfaction resulting from interviews of a panel of experts.
5.2—Vibrations	5.2.1—Evaluations of the internal vibrations (in terms of amplitude and frequency) in predefined running conditions.
5.3—Noise	5.3.1—Maximum noise (in dB) measured in predefined running conditions.
6.1—Cost of raw materials	6.1.1—Estimated cost of the raw materials needed to produce a car.
6.2—Purchase cost of automotive subsystems from third parties	6.2.1—Estimated purchase cost of the automotive subsystems needed to produce a car.
7.1—Manufacturing cost	7.1.1—Estimated average manufacturing cost to produce a car.
7.2—Assembly cost	7.2.1—Estimated average assembly cost to produce a car.
7.3—Logistics cost	7.3.1—Estimated average logistics cost to produce a car.

Regarding the dimension “1. Technical features” and the representation target “1.1—Car mass”, the indicator “1.1.1—Car mass in medium-load conditions” does not seem to be exhaustive: in fact it could also be reasonable to consider the car mass in full-load conditions; in this case, this indicator would have to be modified or completed by new indicators. Additionally, there is no indicator concerning the “external dimensions” of the car (e.g., “length”, “height”, “width”) or other important features, such as the “engine type” (e.g., petrol, diesel, supercharged, electric, hybrid, etc.). As a consequence, new representation targets and relevant indicators could be added (see Table 4.13).

Table 4.13 Additional indicators that complete the representation of the first dimension ("1. Technical features")

Representation target	Indicators
1.1—Car mass	1.1.2—Car mass in full-load conditions.
1.2—Engine characteristics	1.2.2—Petrol, diesel, supercharged, electric, hybrid, etc.
1.4—External dimensions	1.4.1—Length [mm]; 1.4.2—Width [mm]; 1.4.3—Height [mm].

The initial list of indicators is reported on Table 4.12

This "trial-and-error" procedure can be extended to the remaining dimensions and relevant representation targets.

Non-redundancy should be tested analysing each set of indicators. For example, since indicators "5.2—Vibrations" and "5.3—Noise" are potentially correlated, it would be reasonable to classify one of them as redundant and then remove it. In general, potential correlations must be sought among the indicators that represent the same process dimension.

- *Testing single indicators*

The proposed procedure continues considering the remaining properties of single indicators (i.e., *simplicity of use*, *economic impact*, *counter-productivity*, etc.). For example, regarding the indicators "1.4.1—Length", "1.4.2—Width" and "1.4.3—Height", it would be reasonable to consider the accuracy of the measuring instruments.

Economic impact and *simplicity of use* are generally evaluated in relative terms. For example, considering the economic impact of indicator "2.1.1—Maximum engine power", bench tests (in controlled conditions) are generally more practical than road tests.

Testing the *counter-productivity* of indicators can be quite difficult, as all the possible (negative) correlations among indicators should be investigated. For instance, the indicator "2.1.1—Maximum engine power" should not increase too much, in order not to penalize other indicators, such as "2.3.1—Maximum speed along a predefined curve". In fact, excessive engine power may not be transferred completely to the ground, with dangerous consequences for safety (e.g., understeer/oversteer). Additionally, indicator 2.1.1 could be negatively correlated with other ones, such as those related to "noise" and "vibrations" (i.e., 5.2.1 and 5.3.1).

- *Testing derived indicators.*

For each *derived* indicator, for example "4.1.5—Derived indicator summarizing the global pollution level", it is possible to test the properties of *monotony* and *compensation*. Of course, derived indicators that are strictly monotonous with respect to sub-indicators and that satisfy the compensation property would be desirable.

References

Brown, M. G. (1996). *Keeping score: Using the right metrics to drive world-class performance*. New York: Quality Resources, CRC Press.

Caplice, C., & Sheffi, Y. (1994). A review and evaluation of logistics metrics. *The International Journal of Logistics Management, 5*(2), 11–28.

Caplice, C., & Sheffi, Y. (1995). A review and evaluation of logistics performance measurement systems. *The International Journal of Logistics Management, 6*(1), 61–64.

Denton, D. K. (2005). Measuring relevant things. *International Journal of Productivity and Performance Management, 54*(4), 278–287.

Edwards, J. B. (1986). *The use of performance measures*. Montvale, NJ: National Association of Accountants.

Finkelstein, L. (2003). Widely, strongly and weakly defined measurement. *Measurement, 34*(1), 39–48.

Flapper, S. D. P., Fortuin, L., & Stoop, P. P. M. (1996). Toward consistent performance measurement systems. *International Journal of Operations and Production Management, 16*(7), 27–37.

Franceschini, F. (2001). *Dai Prodotti ai Servizi. Le nuove frontiere per la misura della qualità*. Torino: UTET Libreria.

Franceschini, F., & Galetto, M. (2001). A new approach for evaluation of risk priorities of failure modes in FMEA. *International Journal of Production Research, 39*(13), 2991–3002.

Franceschini, F., Galetto, M., Maisano, D., & Mastrogiacomo, L. (2008). Properties of performance indicators in operations management: A reference framework. *International Journal of Productivity and Performance Management, 57*(2), 137–155.

Franceschini, F., Galetto, M., Maisano, D., & Mastrogiacomo, L. (2012). The success-index: An alternative approach to the h-index for evaluating an individual's research output. *Scientometrics, 92*(3), 621–641.

Franceschini, F., Galetto, M., Maisano, D., & Viticchiè, L. (2006a). The condition of uniqueness in manufacturing process representation by performance/quality indicators. *Quality and Reliability Engineering International, 22*(5), 567–580.

Franceschini, F., Galetto, M., & Maisano, D. (2006b). Classification of performance and quality indicators in manufacturing. *International Journal of Services and Operations Management, 2*(3), 294–311.

Franceschini, F., & Maisano, D. (2010). The citation triad: An overview of a scientist's publication output based on Ferrers diagrams. *Journal of Informetrics, 4*(4), 503–511.

Franceschini, F., & Maisano, D. (2017). Critical remarks on the Italian research assessment exercise VQR 2011–2014. *Journal of Informetrics, 11*(2), 337–357.

Galbraith, L., & Greene, T. J. (1995). Manufacturing system performance sensitivity to selection of product design metrics. *Journal of Manufacturing Systems, 14*(2), 71–79.

Galbraith, L., Miller, W. A., & Greene, T. J. (1991). Pull system performance measures: A review of approaches for system design and control. *Production Planning and Control, 2*(1), 24–35.

Gunasekaran, A., & Kobu, B. (2007). Performance measures and metrics in logistics and supply chain management: A review of recent literature (1995–2004) for research and applications. *International Journal of Production Research, 45*(12), 2819–2840.

Hauser, J., & Katz, G. (1998). Metrics: You are what you measure! *European Management Journal, 16*(5), 517–528.

Hazewinkel, M. (2013). *Encyclopaedia of mathematics* (Vol. 2, C). An updated and annotated translation of the soviet "mathematical encyclopaedia". Berlin: Springer Science & Business Media.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences, 102*(46), 16569–16572.

Joint Research Centre-European Commission. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. OECD Publishing. ISBN: 9789264043459.

Juran, J. M. (1988). *Juran on planning for quality*. New York: The Free Press.

Kaplan, R. S., & Norton, D. (1996). *The balanced scorecard*. Cambridge, MA: Harvard Business School Press.

Kaydos, W. (1999). *Operational performance measurement: Increasing total productivity*. Boca Raton, FL: St. Lucie Press.

Kearney, A. T. (1991). *Measuring and improving productivity in the logistics process: Achieving customer satisfaction breakthroughs*. Chicago: Council of Logistics Management.

Keeney, R., & Raiffa, H. (1976). *Decisions with multiple objectives: Preference and value tradeoffs*. New York: Wiley.

Lohman, C., Fortuin, L., & Wouters, M. (2004). Designing a performance measurement system: A case study. *European Journal of Operational Research, 156*, 267–286.

Maskell, B. H. (1991). *Performance measurement for world class manufacturing*. Cambridge, MA: Productivity Press.

Melnyk, S. A., Stewart, D. M., & Swink, M. (2004). Metrics and performance measurement in operations management: Dealing with the metrics maze. *Journal of Operations Management, 22*, 209–217.

Mentzer, J. T., & Konrad, B. P. (1991). An efficiency/effectiveness approach to logistics performance analysis. *Journal of Business Logistics, 12*(1), 33.

Mock, T. J., & Grove, H. D. (1979). *Measurement, accounting and organizational information*. New York: Wiley.

Narens, L. (1996). A theory of ratio magnitude estimation. *Journal of Mathematical Psychology, 40*(2), 109–129.

Narens, L. (2002). A meaningful justification for the representational theory of measurement. *Journal of Mathematical Psychology, 46*(6), 746–768.

Neely, A., Gregory, M., & Platts, K. (1995). Performance measurement system design. *International Journal of Operations and Production Management, 4*, 80–116.

Nevem Workgroup. (1989). *Performance indicators in logistics*. Bedford: IFS Ltd./Springer.

New, C. C., & Szwejczewski, M. (1995). Performance measurement and the focused factory: Empirical evidence. *International Journal of Operations and Production Management, 15*(4), 63–79.

Perrin, B. (1998). Effective use and misuse of performance measurement. *American Journal of Evaluation, 19*(3), 367–379.

Roberts, F. S. (1979). *Measurement theory*. Reading, MA: Addison-Wesley.

Roy, B., & Bouyssou, D. (1993). *Aide Multicritère à la Décision: Méthodes et Cas*. Paris: Economica.

Smith, D. (2000). *The measurement nightmare: How the theory of constraints can resolve conflicting strategies, policies, and measures*. Boca Raton, FL: St. Lucie Press.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*(2684), 677–680.

Stevens, S. S. (1951). Mathematics, measurement and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1–49). New York: Wiley.

Stockburger, D. W. (2016). *Introductory statistics: Concepts, models and applications* (3rd ed.). Retrieved September 2018, from http://psychstat3.missouristate.edu/Documents/IntroBook3/sbk.htm.

U.S. Department of Energy – PBM SIG (Performance-Based Management Special Interest Group). (2012). *Performance-based management handbook of techniques and tools–How to measure performance* (revised). Retrieved September 2018, from http://www.orau.gov/pbm

Winston, J. A. (1999). Performance indicators – Promises unmet: A response to Perrin. *American Journal of Evaluation, 20*(1), 95–99.

Designing a Performance Measurement System

5

Abstract

The present chapter discusses a core problem of quality management for organizations: establishing and maintaining a *performance measurement system*. Flowing from the mission and strategic planning of one organization, a performance measurement system is supposed to include the data to collect, analyze, report and, finally, use to make sound business decisions.

The remainder of this chapter is divided into eight sections. Sections 5.1 and 5.2 present the basic characteristics of an integrated performance measurement system. Section 5.3 discusses some popular approaches to develop performance measurement systems: the *Balanced Scorecard*, the *Critical Few*, the *Performance Dashboards*, and the EFQM model. Description is supported by the use of practical examples. Sections 5.4, 5.5 and 5.6 discuss how to develop/synthesise indicators and maintain an effective performance measurement system. Sections 5.7 and 5.8 deal with the possible misuse of indicators and their impact on organizations.

5.1 The Concept of Performance Measurement System

The concept of *performance measurement* is straightforward: you get what you measure and you cannot manage a process unless you measure it.

In the document *Performance Measurement and Evaluation: Definitions and Relationships* (GAO/GGD-98-26), the U.S. General Accounting Office (2012) provides the following definition: "*Performance measurementis the ongoing monitoring and reporting of program accomplishments, particularly progress towards pre-established goals. It is typically conducted by program or agency management. Performance indicators may address the type or level of program activities conducted (process), the direct products and services delivered by a program (outputs), and/or the results of those products and services (outcomes).*

F. Franceschini et al., *Designing Performance Measurement Systems*, Management for Professionals, https://doi.org/10.1007/978-3-030-01192-5_5

A "program" may be any activity, project, function, or policy that has an identifiable purpose or set of objectives".

Performance indicators are tools to understand, manage and improve the activities of organizations (Eccles 1991). Effective performance indicators allow us to understand:

- How well we are doing;
- If we are meeting our goals;
- If our customers are satisfied;
- If our processes are in control;
- If and where process improvements are necessary.

The result of a performance measurement is a performance indicator, which is generally expressed by a number and a unit of measurement. The number gives a magnitude (how much) and the unit gives a meaning (what). Indicators are always associated with corresponding representation targets. In Chap. 3 we have illustrated the reason why measurements can be considered as special indicators.

As anticipated in Sect. 1.3, most of the indicators of a generic process concern the following aspects:

- *Effectiveness*: a process characteristic indicating the degree to which the process output conforms to requirements ("Are we doing the right things?");
- *Efficiency*: a process characteristic indicating the degree to which the process produces the required output at minimum resource cost ("Are we doing things right?");
- *Customer care*: the degree of satisfaction of process users.

5.1.1 Why Performance Indicators?

Here we briefly introduce some practical reasons for adopting a performance measurement system (Bourne and Bourne 2011):

- A performance measurement system represents a structured approach for focusing on a program's strategic plan, goals and performance;
- Indicators focus on the aspects that deserve more attention for achieving the required output. Indicators provide feedback on progress toward objectives;
- Performance indicators improve *internal* communication (among employees) and *external* communication (among the organization and customers/stakeholders). The emphasis on measuring and improving performance (*results-oriented management*) affects multiple aspects of organizations;
- Performance indicators help justify programs and their cost, i.e., they may demonstrate the potential of a program, supporting the decision-making process.

5.1.2 What Performance Indicators Will Not Tell You

Even though a performance measurement system is a valuable tool to manage and control process development, it will not tell you several things:

- *The cause and effect of outcomes are not easily established.* Outcomes can, and often do, reveal the impact of the program, but without collaborating data, it is difficult to demonstrate that your program was the cause of the outcome(s). The outcomes of a methodology are inevitably affected by many events outside control. Another conditioning element is the time difference between cause and effect.
- *Poor results do not necessarily point to poor execution.* If the performance objectives are not being met, it is obvious that something is wrong, but performance information does not always provide the reason. Instead, it raises a flag requiring investigation. Possible reasons are performance expectations that were unrealistic or changed work priorities.
- *Indicators are only a representation model of a process.* The represented process is not the same as the actual one; it is only a "proxy". So, the level of detail depends on the model.
- *Performance indicators do not ensure compliance with laws and regulations*. Performance indicators do not provide information on adherence to laws and regulations or the effectiveness of internal controls. For example, a new building can be constructed more quickly when safety controls and funding limitations are ignored (see Example 4.18 on Chap. 4). Since compliance and internal controls often have a direct effect on performance, care should be taken when selecting suitable performance indicators. This could be obtained controlling the adherence of activities with respect to laws and regulations.

5.1.3 Problems in Implementing a Performance Measurement System

Brown's quote well synthesizes the possible problems in the construction of a performance measurement system: "*The most common mistake that organizations make is measuring too many variables. The next most common mistake is measuring too few*" (Brown 1996). Precisely, the most common difficulties are:

- *Amassing too much (or too little) data*. Consequently, data may be ignored or used ineffectively.
- *Focusing on short-term indicators*. Most organizations only collect financial and operational data, forgetting to focus on longer-term indicators (cf. Sect. 4.7 on "accessory properties");

- *Collecting inconsistent, conflicting, and unnecessary data* (Flapper et al. 1996; Schmenner and Vollmann 1994). In fact, all indicators should lead to the ultimate success of the organization. An example of conflicting indicators would be measuring reduction of office space per staff, while, at the same time, measuring staff satisfaction regarding the facilities;
- *Indicators may not be linked to the organization's strategic targets*;
- *Inadequate balancing of the organization's performances*. For instance, a restaurant may have perfect kitchen efficiency (in terms of reduction of food waste) by waiting until the food is ordered before cooking it. However, the result of this action may dissatisfy customers because of the long wait (cf. property of *non-couter-productivity*, in Sect. 4.6);
- *Measuring progress too often or not often enough*. There has to be a balance here. Measuring progress too often could result in unnecessary effort and excessive costs, with little or no added value. On the other hand, not measuring progress may lead to ignore potential problems until it is too late to take appropriate action.

5.2 The Construction Process

Generally speaking, processes can be considered like natural organisms evolving over time and influenced by the external environment. The process manager defines targets and the corresponding performance indicators. All interested parties should know process targets, how they contribute to the process success, and the stakeholders' measurable expectations (Neely et al. 1997; Neely 2002).

When establishing a performance measurement system, three basic aspects should be considered, as described in the following subsections:

- strategic plan;
- key sub-processes;
- stakeholder needs.

5.2.1 The Strategic Plan

Strategic plan sets the foundation for effective performance measurement systems. Performance measurement systems that focus on the wrong set of indicators can actually undermine an organization's strategic mission by perpetuating short-sighted business practices. For this reason, it is appropriate to discuss the critical elements of a strategic plan and review its compatibility with the integrated performance measurement system. A well-developed strategic plan should contain the basic information for constructing an integrated performance measurement system, as shown in Table 5.1.

Table 5.1 Strategic plan element and performance measurement attributes

Strategic-plan determinant	Description
Strategic Goal	It determines the final mission of the process of interest.
Objective	It describes the strategic activities that are required to accomplish the goal.
Strategy	It defines strategic long-term requirements that link to objectives. E.g., requirements concerning performance targets, strategic dates, etc.
Tactical Plan	It identifies the short-term requirements that link to strategy. E.g., requirements concerning cost, time, milestone, quality, safety attributes, etc.

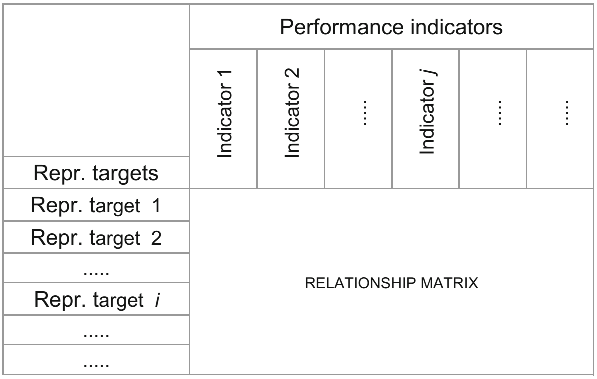

Fig. 5.1 Example of performance-indicator mapping. Representation targets related to the strategic plan are reported at the left of the Relationship Matrix (rows). Performance indicators are reported at the top of the Relationship Matrix (columns)

Having defined a strategic plan, we should determine a suitable set of indicators and focus on the needed information. The objective is to find out which indicators should be monitored and mantained, and who are the owner(s) and data customer(s). Answering the following five questions should provide enough information for this step:

- What information is being reported?
- Who is responsible for collecting and reporting data on the process performance?
- When and how often are performance indicators reported?
- How is the information reported?
- To whom is the performance indicator reported?

Performance indicators and strategic plan can be linked by using a spreadsheet or a table, as the one used for the *Quality Function Deployment* methodology (Franceschini 2002). Figure 5.1 shows an example of mapping. Representation targets are organized hierarchically and reported in the rows of the so-called *Relationship Matrix*. Performance indicators are reported in the columns of the Relationship Matrix.

Key:

Ø: no relationship
Δ: weak relationship
O: medium relationship
◉: strong relationship

Represent. targets	Importance	Indicators: Routing effectiveness	Accuracy in responses	Uniformity of responses (by different operators)	Request-implementation time	Competence perception	Percentage of answered calls	Courtesy of responses	Confidential-data security	Number of active lines
Reliability	5	◉	O	◉	O	Δ			O	
Responsiveness	5				◉	Δ				
Competence	4	◉	◉	◉		◉				
Access	4						◉			O
Courtesy	3			Δ				◉		
Communication	3		Δ	◉		Δ		O		
Credibility	3	Δ		O		O				
Security	5	Δ							◉	
Understanding/ Knowing the customer	4	O	Δ	Δ		Δ		O		
Tangibles	3									◉
Current values		93%	H	H	22 min	VG	98%	M	3%	3

Fig. 5.2 Relationship Matrix related to a help desk service. Relations between representation targets and performance indicators are represented by specific symbols (DISPEA 2005)

Representation targets may usually "impact" on different performance indicators and *vice versa*. The Relationship Matrix, $\mathbf{R} \in \mathfrak{R}^{m,\ n}$ (being m the number of targets and n the number of indicators), represents the relations among representation targets and performance indicators.

These relations are represented by specific symbols: a triangle for *weak* relationships, a circle for *medium* relationships, and two concentric circles for *strong* relationships (Fig. 5.2). Empty cells depict the absence of relationships among representation targets and performance indicators. This methodology makes it possible to transform representation targets into a structured set of control actions. For the purpose of example, Fig. 5.2 shows the Relationship Matrix related to the help desk service for the ICT (Information Communication Technology) function of an important Italian broadcasting company (DISPEA 2005).

Let us focus the attention on the current/desired values of performance indicators (see the bottom of Fig. 5.2), which can be used to plan future improvement actions for the help-desk service.

Considering the example in Fig. 5.2, the representation targets of the service are selected using the model suggested by Parasuraman, Zeithaml and Berry (PZB model)—i.e., one of the most popular models in the scientific literature for evaluating the quality of services. The model identifies 10 key representation targets

Table 5.2 Determinants for the service quality, according to the PZB model (Parasuraman et al. 1985). With permission

Determinant	Description
Reliability	Ability to perform the promised service in a dependable and accurate manner. The service is performed correctly on the first occasion, the accounting is correct, records are up to date and schedules are kept.
Responsiveness	Readiness and willingness of employees to help customers by providing prompt timely services, for example, mailing a transaction slip immediately or setting up appointments quickly.
Competence	Possession of the required skills and knowledge to perform the service. For example, there may be competence in the knowledge and skill of contact personnel, knowledge and skill of operational support personnel and research capabilities of the organization.
Access	Approachability and ease of contact. For example, convenient office operation hours and locations.
Courtesy	Consideration for the customer's property and a clean and neat appearance of contact personnel, manifesting as politeness, respect, and friendliness.
Communication	It means both informing customers in a language they are able to understand and also listening to customers. An organization may need to adjust its language for the varying needs of its customers. Information might include, for example, explanation of the service and its cost, the relationship between services and costs and assurances as to the way any problems are effectively managed.
Credibility	It includes factors such as trustworthiness, belief and honesty. It involves having the customer's best interests at prime position. It may be influenced by company name, company reputation and the personal characteristics of the contact personnel.
Security	It enables the customer to feel free from danger, risk or doubt including physical safety, financial security and confidentiality.
Knowing the customer	It means making an effort to understand the customer's individual needs, providing individualized attention, recognizing the customer when they arrive and so on. This in turn helps to delight the customers by rising above their expectations.
Tangibles	They are the physical evidence of the service, for instance, the appearance of the physical facilities, tools and equipment used to provide the service; the appearance of personnel and communication materials and the presence of other customers in the service facility.

for service quality, called "determinants" (see Table 5.2) (Parasuraman et al. 1985; Franceschini 2001; Franceschini and Mastrogiacomo 2018).

If necessary, each determinant (or first-level item) may be "exploded" into more detailed (second-level) items, depending on the examined process (see the example in Table 5.3). Typical techniques to identify the service targets are personal interviews and focus groups (Franceschini 2002).

Since the representation targets of the process may have a different importance, they can be ranked. A classical method to obtain a ranking is to associate each

Table 5.3 "Explosion" of two determinants (i.e., "reliability" and "responsiveness", from Table 5.2) into multiple second-level items (DISPEA 2005)

Determinant	Second level item
Reliability	Accuracy in service pricing.
	Service punctuality.
	Prompt feedback to customers.
Responsiveness	Promptness in fixing appointments.
	Promptness in forwarding the documentation to customers.
	Skill and knowledge of front-office operators.

Table 5.4 List of indicators determined for a help desk service (see Fig. 5.2) (DISPEA 2005)

Indicator	Definition	Measuring scale
Routing effectiveness	Percentage of calls switched less than three times.	%
Accuracy in responses	Ability of resolving calls accurately, exhaustively and consistently with protocols.	High/Medium/Low
Uniformity of responses	Degree of uniformity among operators.	High/Medium/Low
Implementation time	Average time before implementing a customer request.	Minutes
Competence perception	Customer perception of the competence of operators in providing exhaustive responses.	Poor/Sufficient/Good/ Very good/Excellent
Percentage of answered calls	Ratio between the number of answered calls and the total number of calls received by the help desk.	%
Courtesy of responses	Operators' courtesy when answering the customer calls.	High/Medium/Low
Confidential-data security	Portion of customer complaints concerning the loss of confidential data.	%
Number of active lines	Number of telephone lines available.	Number

representation target with an importance, e.g., from 1 (very low importance) to 5 (very high importance); see Fig. 5.2.

Representation targets are then translated into indicators that are supposed to represent them exhaustively. In practice, it is possible to construct a Relationship Matrix, as exemplified in Fig. 5.2. For each indicator, it is necessary to provide a clear definition of (i) the measurement scale/unit and (ii) the data-collection procedure. Table 5.4 illustrates the indicators related to the example in Table 5.2.

5.2.2 Key Sub-processes

Processes represent the implementation of a strategic plan. When the complexity of a process tends to be high, it may be approprate to decompose a process into (key) sub-processes, which are organized hierarchycally, depending on their impact on targets.

This can done using the so-called "process maps", i.e., some graphical tools containing the relevant (qualitative and quantitative) information on the process of interest. Process maps provide a visual representation of the activities, interfaces, flows of information, and responsibilities of process actors; in other words, a detailed "snapshot" of the organization.

The methodology is structured into the following stages:

1. Preliminary analysis of processes;
2. Drafting process maps;
3. Analysing process maps.

Preliminary Analysis of Processes

The purpose of this stage is to obtain a general picture of the process. The fundamental steps are:

- Identifying the core-activities of the organization of interest;
- Identifying the interface(s) between the organization and customers;
- Identifying the interface(s) between the organization and the remaining stakeholders;
- Identifying the typical data to manage and the potential problems/criticalities.

This stage is very important to (re)structure the current process organization.

Drafting Process Maps

A generic process can decomposed into further sub-process. This decomposition may be performed at different levels, as illustrated in Fig. 5.3.

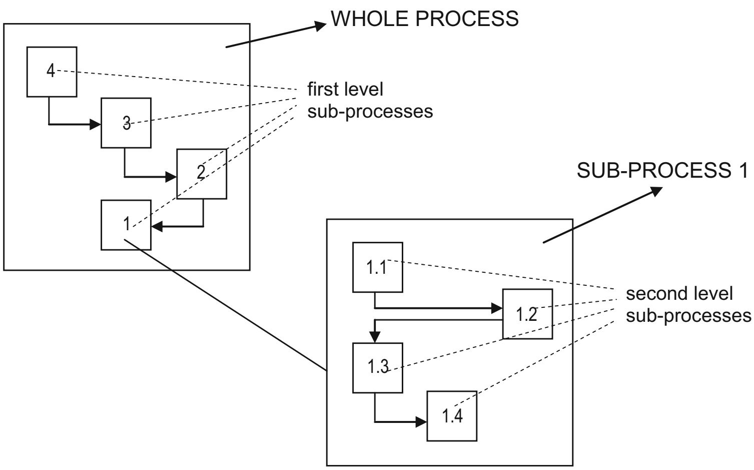

Fig. 5.3 Example of process decomposition into sub-processes

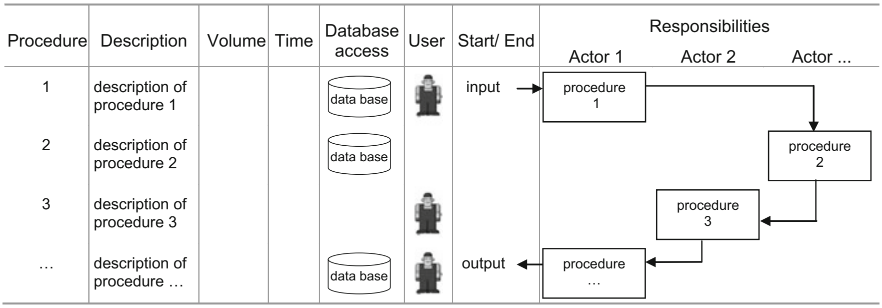

Fig. 5.4 General scheme of a process map

The level of detail depends on the complexity of the whole process. Additionally, the number of levels should not be higher than 4 or 5, in order not to complicate the analysis. Figure 5.4 shows an example of process map.

The first action in the construction of a process map, taking the following features into account:

- Process *input*;
- Procedures/activities (represented by blocks in the process map);
- Process *output* (results).

Then, responsibilities of actors (i.e., "who does what") are defined, representing the information flow graphically.

Process maps may also include information on (i) data at the interface between organization and customers, (ii) data shared with suppliers, (iii) time required by the various activities, etc.

Analysing Process Maps

The last stage of the methodology is the analysis of process maps, with the purpose of determining the efficiency/effectiveness of the process, and *how*, *where* and *when* quality monitoring should be performed. In addition, the weaknesses of the process should be identified and fixed.

A "vertical reading" of process maps allows to identify the process activities related to the single actors. Moreover, a virtual superimposition of several process maps allows to identify the workloads and so-called "vicious circles" related to the various activities.

5.2.3 Stakeholder Needs

Stakeholders are those people with a stake in the future success of an organization. It is important to have a clear idea of stakeholders' needs and expectations. A strategy to systematically understand what stakeholders want and expect does not yet exist

(Atkinson 1997). Regarding customers, surveys or focus groups are often used; regarding employees, surveys, focus groups or discussions are often used too.

Developing performance indicators related to stakeholders may help in:

- Understanding whether tactical plans are being met. E.g., those concerning customer satisfaction and employee commitment;
- Testing the presumed cause-and-effect relationships between performance indicators and strategies. E.g., does higher quality always result into increased sales?

In many organizations, leadership commitment to the development and use of performance indicators is very important. Four specific ways to encourage this commitment are (Thomson and Varley 1997):

- *Delegate responsibilities and empower employees*. Involvement creates ownership and encourage loyalty, commitment and accountability of employees.
- *Develop good communication processes*. A good communication process provides a critical link between the tasks of employees and the strategic-plan indicators.
- *Always seek feedback*. Managers need to know what employees think about their activities, especially if they are not aligned with the strategic direction of the organization. Doing so creates accountability for both employees and senior management.
- *Definition of responsibilities*. Each performance indicator needs someone that is responsible for it. In addition, employees need to know how indicator(s) relate to the success/failure of the organization.

5.2.4 Vertical and Horizontal Integration of Performance Indicators

Performance indicators need to be integrated in two directions: vertically and horizontally. *Vertical* integration may encourage employees to focus on the organization's strategic objectives. On the other hand, *horizontal* integration encourages the optimization of work flow across all process and organizational boundaries. Figure 5.5 provides an example of "deployment" of indicators at different organizational levels.

Vertically integrated performance indicators include several requirements:

- Determining target values;
- Integrating process indicators and results;
- Defining level-by-level responsibilities;
- Coordinating data collection and data analysis.

From the customer viewpoint, organizations are often perceived as perfectly structured bodies, with no boundaries between activities and/or functions. However,

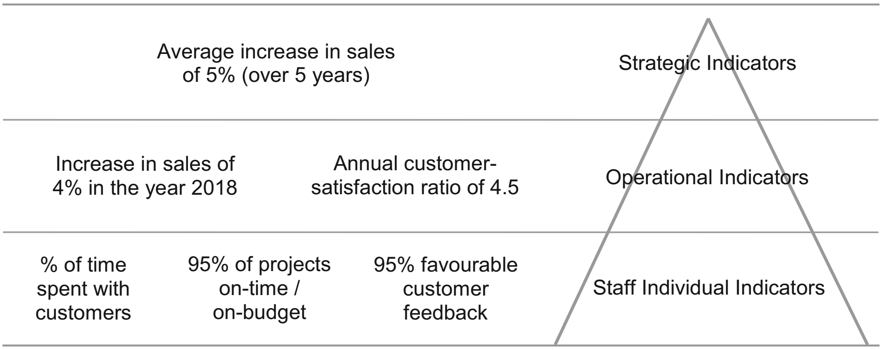

Fig. 5.5 Example of "deployment" of indicators at different organizational levels (U.S. Department of Energy – PBM SIG 2012). With permission

activities and functions must be coordinated and aligned with each other, sharing the relevant information: this is the essence of the concept of *horizontal integration*.

5.3 A Review of the Major Reference Models

When developing a performance measurement system, it is convenient to follow a suitable reference model, especially when doing it for the first time. The scientific literature includes several popular approaches; the following subsections provide a short description of some of them.

5.3.1 The Concept of "Balancing"

The concept of *balancing* performance indicators was introduced in 1992, when Robert Kaplan and David Norton—from Harvard University—developed the *Balanced Scorecard* methodology. In a nutshell, the basic idea is to translate the business objectives of an organization into a *critical set* of indicators, according to the major *dimensions* (or perspectives). Additionally, these performance indicators should be balanced, trying to minimize negative competition between individuals/functions.

This concept is the starting point of several operational approaches:

- the Balanced Scorecard method;
- the Critical Few method;
- the Performance Dashboards;
- the EFQM (European Foundation for Quality Management) model, which is very close to the Malcolm Baldridge National Quality Award (American) model.

These approaches will be presented and discussed in the following sections.

5.3.2 The "Balanced Scorecard" Method

In 1992, Robert Kaplan and David Norton introduced the *Balanced Scorecard* concept as a way of motivating and measuring an organization's performance (Kaplan and Norton 1992, 2008).

The concept takes a systematic approach in assessing internal results, while probing the external environment. It focuses as much on the process of arriving at "successful" results, as on the results themselves.

Indicators that make one *dimension* look good while deflating another are avoided, thus minimizing negative competition between individuals and functions. This framework is intended for top managers willing to obtain a quick and comprehensive assessment of their organization in a single report. The *Balanced Scorecard* encourages the reduction of the amount of indicators to a "vital few", in order to understand whether results in one area are being achieved at the expense of another area.

The method considers four interconnected dimensions:

- *Financial*. How do we look to our stakeholders?
- *Customer*. How well do we satisfy our internal and external customer's needs?
- *Internal Business Process*. How well do we perform at key internal business (sub) processes?
- *Learning and Growth*. Are we able to sustain innovation, change, and continuous improvement?

A visual representation of these dimensions is provided in Fig. 5.6. The *Balanced Scorecard* may help managers to obtain an exhaustive multi-perspective view of their organization (cfr. Sect. 4.6.2). Each dimension is related to specific performance targets. In this framework, (1) customer satisfaction drives financial success, (2) effective and efficient processes ensure high customer satisfaction, and (3) continuous improvement enhances operational performance. Figure 5.6 also shows that each dimension may also influence (some of) the other ones.

Performance indicators related to each of these dimensions can be found answering the following questions:

- *Financial*. What are the strategic financial objectives?
- *Customer*. What do we have to do for our customers in order to achieve financial success?
- *Internal Business Process*. Which of our business processes mostly impact on customer satisfaction?
- *Learning and Growth*. What are the improvements needed to obtain sound business processes and satisfied customers?

The concept of balance consists in putting the right emphasis on all the important dimensions of the process/organization of interest. For example, performance

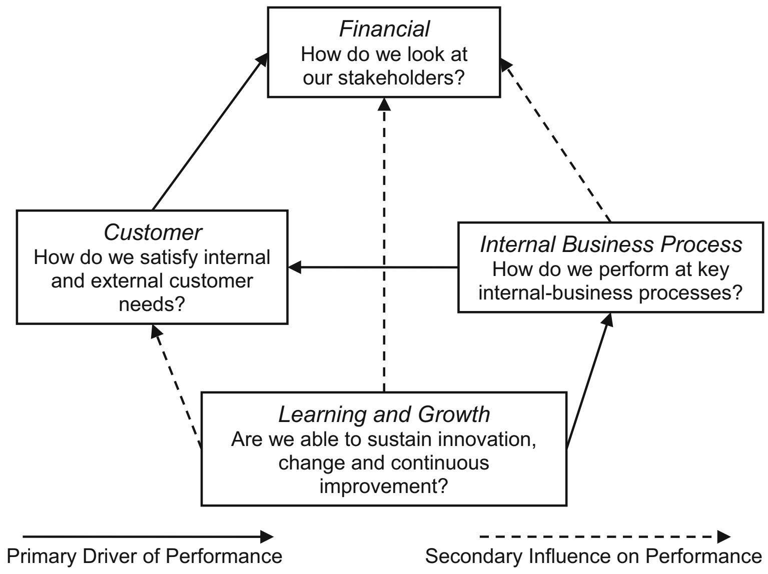

Fig. 5.6 The four dimensions of the *Balanced Scorecard* model (Kaplan and Norton 1992). With permission

indicators that consider the economic/financial dimension only (neglecting the others) should be avoided (see Sect. 1.4.6).

5.3.3 The "Critical Few" method

Managing too many indicators may have different drawbacks, such as (1) losing sight of their impact, (2) distracting management's focus from the most critical indicators for success, and (3) neglecting possible correlations between indicators. Consequently, the number of indicators should be as small as possible, as long as the exhaustiveness of representation is preserved: *pluritas non est potenda sine necessitate* (Thorborn 1918). The process of simplifying and "distilling" a large number of performance indicators into a "critical few" should be viewed as part of the performance measurement process itself, which may improve the understanding of strategic objectives.

The selection of a critical-few set of performance indicators requires a balance between internal and external requirements, as well as financial and non-financial ones. Although there is no a "magical" number of critical-few indicators, some guidelines suggest a number included between 3 and 15 for each organizational level (U.S. Department of Energy – PBM SIG 2012).

Likewise the Balanced Scorecard, the Critical Few framework includes several strategically focused business dimensions, identifying performance objectives and indicators for each dimension. For example, let us consider the 35 indicators selected for monitoring a call center in Table 5.5 (DISPEA 2004). It is clear that considering them all could be difficult and even impractical. The Critical Few method makes it possible to reduce the number of indicators, without compromising the process control. Section 5.4 will present some practical techniques to this purpose.

5.3.4 Performance Dashboard

A performance dashboard is a practical tool that synthesizes the performance level of a process. In Europe, several organizations have developed the *Tableau de Bord*, i.e., a dashboard of key performance indicators, which synthesize the most relevant factors for success.

The basic idea is that leading an organization is like driving a car. After all, there are not so many gauges on the dashboard: the most relevant ones are “level of fuel”, “water level” and “emergency lights”. The driver’s primary focus is to move the car safely in one direction while watching for obstacles on the road (including other cars!).

This is exactly what a good “driver” in an organization should be doing. A balanced set of performance indicators is like gauges on the dashboard of a car; destination is the mission. Each gauge represents an aggregated (or derived) indicator, summarizing the performance of a relevant part of the process of interest. For example, the temperature gauge could represent customer satisfaction and aggregate several sub-indicators, such as “number of customer complaints”, “organization reputation”, etc.

For the purpose of example, the Balanced Scorecard might be presented as a performance dashboard, as shown in Fig. 5.7.

Designing a “good” dashboard is essential to avoid that indicators are inexhaustive, dispersive and not adequately representative of different areas and responsibilities of the organization. When defining the data to be collected, one should clearly understand the purpose of the representation and the information capacity of indicators. To simplify the design, it is possible to construct different dashboards representing different areas of the organization. Each dashboard should be “modular”, allowing—if necessary—a decomposition/specialization of the relevant data (Lohman et al. 2004).

Example 5.1 The University of California manages three national laboratories for the Department of Energy: *Lawrence Berkeley*, *Lawrence Livermore*, and *Los Alamos*. The overall performance of these laboratories is monitored considering several indicators, related to ten administrative and operational functions (University of California 2017):

Table 5.5 List of indicators selected for monitoring a call center

Target	Indicator	Evaluation scale
Reliability	1.1 Accuracy in responses	High/Medium/Low
	1.2 Uniformity of responses (by different operators)	Yes/No
	1.3 "Routing" (95% of the calls)	No. of switches
Responsiveness	2.1 Time to forward a request to back-office (95% of the calls)	Minute
	2.2 Percentage of calls resolved in 1 day, after being forwarded to the back-office	High/Medium/Low
	2.3 Customer perception of the service responsiveness	P/S/G/VG/E
Competence	3.1 Training of operators	High/Medium/Low
	3.2 Average experience of operators (no. of years)	Number
	3.3 Customer perception of the operator competence	P/S/G/VG/E
Access	4.1 Visibility of the call-center phone number	P/S/G/VG/E
	4.2 Average cost of the call	P/S/G/VG/E
	4.3 Average opening time	P/S/G/VG/E
	4.4 Possibility to contact a previously-contacted operator	P/S/G/VG/E
	4.5 Daily number of received calls	Number
	4.6 Daily percentage of aborted calls	High/Medium/Low
	4.7 Percentage of answered calls in 1 day	High/Medium/Low
	4.8 Average time between service access and operator's answer (95% of the calls)	Minute
	4.9 Daily percentage of calls forwarded to back-office	High/Medium/Low
	4.10 Daily percentage of calls sent (from back-office) back to operators	High/Medium/Low
	4.11 Daily total call time	Minute
	4.12 Average call time (95% of the calls)	Minute
	4.13 Daily percentage of queued calls	High/Medium/Low
	4.14 Maximum queue time in 1 day	Minute
Courtesy	5.1 Average operator's courtesy in opening the call	P/S/G/VG/E
	5.2 Average operator's courtesy in managing the call	P/S/G/VG/E
	5.3 Average operator's courtesy in closing the call	P/S/G/VG/E
Communication	6.1 Response clarity	P/S/G/VG/E
	6.2 Response accuracy	P/S/G/VG/E
	6.3 Response personalization	P/S/G/VG/E
Credibility	7.1 Trust in the operator	P/S/G/VG/E
Security	8.1 Information on the cost of the service	Yes/No
Understanding/knowing the customer	9.1 Data collected from customer requests (percentage)	%
	9.2 Response understanding	P/S/G/VG/E
Tangibles	10.1 Total number of phone lines	Number
	10.2 Existence of tools for managing queued calls	Yes/No

P/S/G/VG/E stand for poor/sufficient/good/very good/excellent respectively (DISPEA 2004)

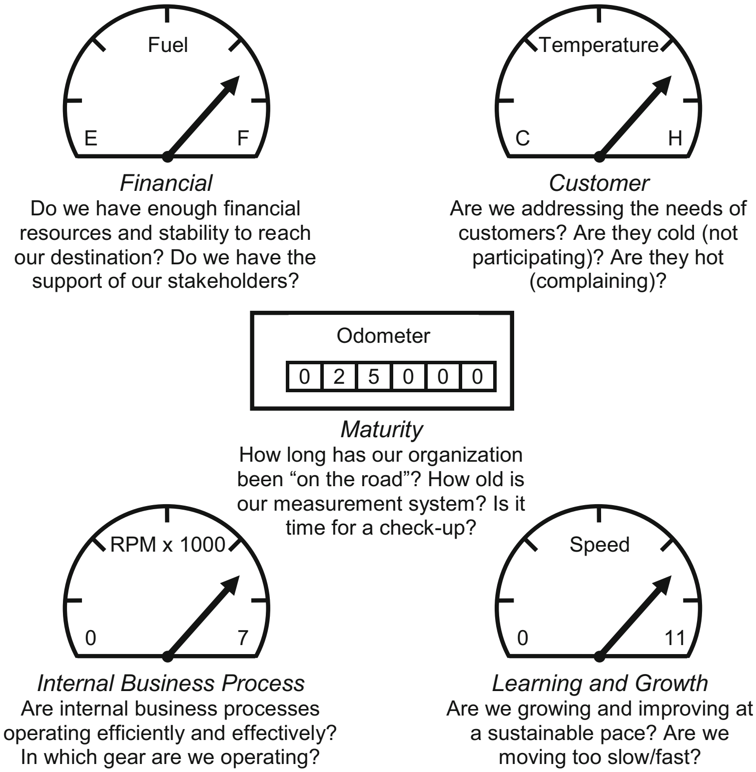

Fig. 5.7 Synthetic representation of the Balanced Scorecard as a performance dashboard (U.S. Department of Energy – PBM SIG 2012). With permission

1. Laboratory management;
2. Environmental restoration and waste management;
3. Environment, safety and health;
4. Facilities/project management;
5. Safeguards and security;
6. Finance;
7. Human resources;
8. Information management;
9. Procurement;
10. Property management.

Each function receives an annual score and the resulting scores are combined into an overall score, which depicts the administrative and operational performance of laboratories. The performance scale is:

Table 5.6 Houghton College's (NY) performance dashboard (Houghton College 2004). With permission

Dashboard indicator	Definition
1. Income stability	Excessive tuition dependence increases volatility, particularly during economic recession and times of demographic change and uncertainty. The income stability dimension focuses on tuition dependency. Its measurement is gross tuition and fees as a percentage of gross Education and General (E&G) revenue.
2. Commitment to academic excellence	Generally, to the extent that we are able to preserve a significant portion of our budget for instruction, we are investing in academic excellence today and in the future. This dimension focuses on instructional expenses. Its measurement is instructional expenses as a percentage of net expenditures.
3. Stewardship	An operating excess generally will mean that we are meeting our budgetary goals and living within our means. The stewardship dimension focuses on financial surplus. Its measurement is total current fund revenues less total current fund expenditures.
4. Competitiveness/ selectivity	These two separate measures are highly interrelated. While the first is a widely used measure of selectivity, the second is a qualitative measure of admissions "yield," an important indication of Houghton's attractiveness. Together they suggest how much flexibility we have to control the quality and composition of our student body. This dimension focuses on selectivity and academic quality. Selectivity is measured in terms of the percentage of applicants accepted as freshmen. Academic quality is measured in terms of the percentage of freshmen who graduated in the top 10 percent of their high school class.
5. Productivity	While our overall ratio of students to faculty may mask significant variability among programs and departments, this indicator is the starting point for assessing faculty workload and productivity. The second indicator, while again tied to the number of students, provides a measure of our administrative productivity. The productivity dimension focuses on faculty and administrative workload. It measures full-time students per full-time faculty member and full-time equivalent students per full-time equivalent staff member.
6. Affordability	The policy of tuition discounting may be justified as long as net tuition (i.e., gross tuition revenue, institutionally-funded aid) continues to grow. This indicator should be considered in light of institutional selectivity, focusing on student aid. It is measured in terms of college-funded student financial aid as a percentage of tuition and fees.
7. Mission and program mix	The proportion of employees who are faculty reflects the college's mission and program mix, as well as its choices about the division of labour between faculty and staff. Precisely, this indicator corresponds to the percentage of full-time employees who are faculty.
8. Facility maintenance	Deferred maintenance is a growing concern for many colleges, whose capital assets are deteriorating as scarce funds are diverted to academic and other priorities that seem to be more pressing. The lower this number, the better. The facility maintenance dimension focuses on facility maintenance backlog. It is measured in terms of the estimated maintenance backlog as a percentage of the total replacement value of the plant.

(continued)

Table 5.6 (continued)

Dashboard indicator	Definition
9. Alumni support	Alumni giving is a significant source of institutional support and an important barometer for constituent opinion about institutional performance. This dimension focuses on alumni contributions. Its measurement is the percentage of alumni who have given at any time during the past year.

– Outstanding (90–100%);
– Excellent (80–89%);
– Good (70–79%);
– Marginal (60–69%);
– Unsatisfactory (<60%).

It can be noticed that the scale is unbalanced, as four (out of five) levels are related to a performance higher than 60%. What is the rationale behind this choice?

Example 5.2 Houghton College (NY) established a dashboard of nine synthetic indicators to monitor the overall performance of the institution (Houghton College 2004). These indicators are defined in Table 5.6.

5.3.5 The EFQM (European Foundation for Quality Management) Model

The *European Foundation for Quality Management* (EFQM) is a non-profit organization, which was created in 1988 with the mission of being the driving force for *Sustainable Excellence* in quality management in Europe (EFQM 2013; APQI 2018).

The EFQM model can be used to assess the progress of an organization towards excellence, independently of the organization type, size, structure, and maturity. The model is a nonprescriptive framework, which recognises several possible approaches to achieving sustainable excellence. It is based on the premise that Excellence depends on the capacity of conciliating the different exigencies and interests of stakeholders'.

The model is based on nine criteria (dimensions). Five of these criteria are classified as "Enablers" and four as "Results". The "Enabler" criteria cover what an organization does; the "Result" criteria cover what an organization achieves. Feedback from "Results" help to improve "Enablers".

The EFQM Model is based on the premise that excellent results—with respect to Performance, Customers, People and Society—are achieved through Leadership driving Policy and Strategy, that is delivered through People Partnerships and Resources, and Processes.

Consistently with what is reported in (EFQM 2013), each criterion refers to a specific examined area and has a weight (percentage), which is used to determine the final score. The weights reported in Fig. 5.8 were defined in 2013 and represent the

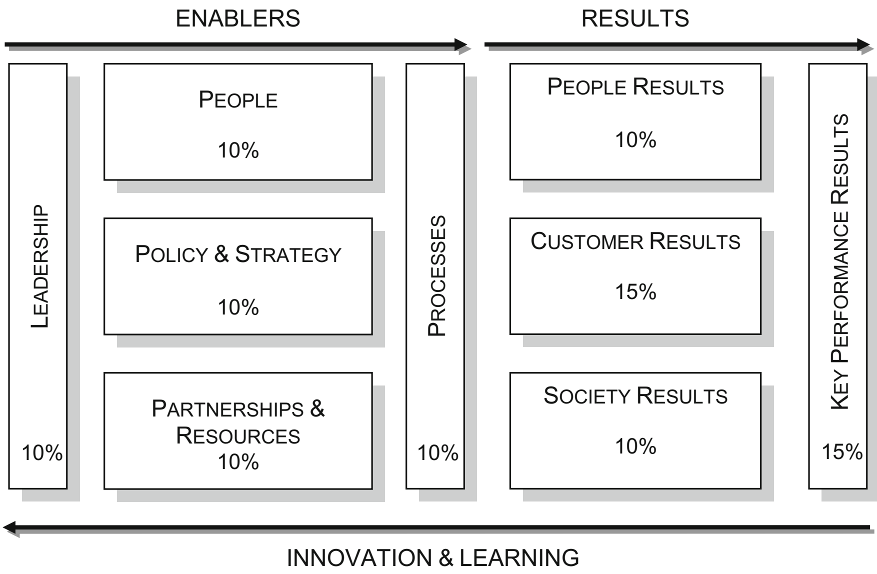

Fig. 5.8 Scheme of the EFQM model. Nine boxes represent the nine criteria of the model: five of them are classified as "Enablers" and the remaining four as "Results". "Enabler" criteria are concerned with what an organization does; "Results" criteria are concerned with what an organization achieves. Each criterion has a corresponding weight (percentage), which is used for determining the final score (EFQM 2013; APQI 2018)

result of an extensive debate among several European organizations. Organizations can use the suggested weights, but they may also adjust them, case by case. The definitions of the criteria, taken from (EFQM 2013), are given below.

Criterion 1—Leadership

Excellent leaders develop and facilitate the achievement of the mission and vision. They develop organizational values and systems required for sustainable success and implement these via their actions and behaviours. During periods of change they retain a constancy of purpose. Criterion 1 can be divided in the following sub-criteria.

1.a Leaders develop the mission, vision, values and ethics and are role models of a culture of Excellence.
1.b Leaders are personally involved in ensuring the organization's management system is developed, implemented and continuously improved.
1.c Leaders interact with customers, partners and representatives of society.
1.d Leaders reinforce a culture of Excellence with the organization's people.
1.e Leaders ensure that the organization is flexible and manages change effectively.

Criterion 2—Policy and Strategy

Excellent organizations implement their mission and vision by developing a stakeholder-focused strategy that takes account of the market and sector in which they operate. Policies, plans, objectives, and processes are developed and deployed to deliver the strategy. Criterion 2 can be divided into the following sub-criteria.

2.a Policy and Strategy are based on the present and future needs and expectations of stakeholders.
2.b Policy and Strategy are based on information from performance indicator, research, learning and external related activities.
2.c Policy and Strategy are developed, reviewed and updated.
2.d Policy and Strategy are communicated and deployed through a framework of key processes.

Criterion 3—People
Excellent organizations manage, develop and release the full potential of their people at an individual, team-based and organizational level. They promote fairness and equality, and involve/empower their people. They care for, communicate, reward and recognise, in a way that motivates staff and builds commitment to using their skills and knowledge for the benefit of the organization. Criterion 3 can be divided into the following sub-criteria.

3.a People resources are planned, managed and improved.
3.b People's knowledge and competencies are identified, developed and sustained.
3.c People are involved and empowered.
3.d People and the organization have a dialogue.
3.e People are rewarded, recognised and cared for.

Criterion 4—Partnership and Resources
Excellent organizations plan and manage external partnerships, suppliers and internal resources, in order to support policy and strategy, and the effective operation of processes. Criterion 4 can be divided into the following sub-criteria.

4.a External partnerships are managed.
4.b Finances are managed.
4.c Buildings, equipment and materials are managed.
4.d Technology is managed.
4.e Information and knowledge are managed

Criterion 5—Processes
Excellent organizations design, manage and improve processes in order to fully satisfy and generate increasing value for customers and other stakeholders. Criterion 5 can be divided into the following sub-criteria.

5.a Processes are systematically designed and managed.
5.b Processes are improved, as needed, using innovation in order to fully satisfy and generate increasing value for customers and other stakeholders.
5.c Products and services are designed and developed based on customer needs and expectations.

5.d Products and services are produced, delivered and serviced.
5.e Customer relationships are managed and enhanced.

Criterion 6—Customer Results
Excellent organizations comprehensively measure and achieve outstanding results with respect to their customers. Criterion 6 can be decomposed into the following sub-criteria.

6.a Perception indicators.
6.b Performance indicators.

Criterion 7—People Results
Excellent organizations comprehensively measure and achieve outstanding results with respect to their people. Criterion 7 can be divided into the following sub-criteria.

7.a Perception indicators.
7.b Performance indicators.

Criterion 8—Society Results
Excellent organizations comprehensively measure and achieve outstanding results with respect to society. Criterion 8 can be divided into the following sub-criteria.

8.a Perception indicators.
8.b Performance indicators.

Criterion 9—Key Performance Results
Excellent organizations comprehensively measure and achieve outstanding results with respect to the key elements of their policy and strategy. Criterion 9 can be divided into the following sub-criteria.

9.a Key performance outcomes.
9.b Key performance indicators.

How to Use Model

The EFQM model is one of the most widely used in Europe and represents the starting point to apply for the European Quality Award. Furthermore, it can be used for many other purposes: for internal self-assessments, to drive future improvements or benchmarking among organizations, etc. The following tools are provided: *Pathfinder Card* and *Radar Scoring Matrix*.

The Pathfinder Card is a self-assessment tool to identify opportunities for improvement and to help build improvement plans.

Tables 5.7 and 5.8 report a sequence of questions that can be used to assess the plausibility of Enablers and Results, respectively. Improvement actions are supposed to be focused on the weaker points.

Table 5.7 Pathfinder Card: Questions for the organization's self-assessment of the Enablers of the EFQM model (Fig. 5.8) (EFQM 2013; APQI 2018). With permission

Enablers section		
Approach	Deployment	Assessment and review
• Does the approach have a clear rationale? • Are processes well defined and developed? • Does the approach focus on stakeholder needs? • Does the approach support policy and strategy? • Is the approach related to other suitable approaches? • Is the approach sustainable? • Is the approach innovative? • Is the approach flexible? • Is the approach measurable?	• Is the approach applied to all the relevant areas of the organization? • Is the approach implemented in a systematic and structured way? • Does the implemented approach fulfil the expected results? • Do stakeholders accept the implemented approach? • Is the approach deployment measurable?	• Is the effectiveness of the approach evaluated periodically? • Are good practices and improvement opportunities identified and shared regularly? • Is our approach comparable with that of competitors (e.g., the "best in class")? • Is the output information used to identify, prioritise, plan and implement improvements?

Table 5.8 Pathfinder Card: Questions for the self-assessment of the Results of the EFQM model (Fig. 5.8) (EFQM 2013; APQI 2018). With permission

Results section
• Do results concern all the stakeholders? • Do results consider the effect of different approaches (and their own deployments) through appropriate performance and perception indicators? • Do results show positive trends and/or determine good performance? If so, how long has this been going on? • Are results consistent with targets? If so, have targets been achieved? • Are our results comparable with those of external organizations (competitors)? • Do results compare well with industry averages or acknowledged "best in class"? • Are results affected by the proposed approach? • Can results be measured (at present time and in the future)? • Do results address all relevant areas?

The RADAR Scoring Matrix is used to assess the Results criteria for the European Quality Award (i.e., criteria 6–9 of the EFQM Model). This method can be adopted by organizations for benchmarking or self-assessment. The final score is a weighted mean of the scores related to the single criteria.

The five Enablers criteria and four Results criteria have a 50% aggregate weight, showing that actions and results are equally important. It can be noticed that, while weights of Enablers criteria are equally distributed, those of Results criteria are not; precisely:

- 6a takes 75% of the points allocated to criterion 6 and 6b takes 25% of the points allocated to criterion 6.
- 7a takes 75% of the points allocated to criterion 7 and 7b takes 25% of the points allocated to criterion 7.
- 8a takes 25% of the points allocated to criterion 8 and 8b takes 75% of the points allocated to criterion 8.

Tables 5.9 and 5.10 show the RADAR Scoring Matrix, with reference to Enablers and Results respectively.

The model determines a total score expressed on a 0-to-1000-point scale. The score allocation reflects the relative weights of the EFQM model (see Fig. 5.8); the Scoring table in Fig. 5.9 supports this calculation.

EFQM model is not the only way to evaluate the performance of organizations. A similar model is the *Malcom Baldrige National Quality Award* (MBNQA), which was established by the United States Congress in 1988, with the aim of addressing the activities of organizations towards the implementation of *Total Quality Management* (TQM) (Juran 2016).

The EFQM model has several other potential uses. For example, it can be used to identify areas for improvements towards excellence or may be helpful to drive internal improvement, independently from the size and field of one organization. On the other hand, the model presents some weaknesses; for example, the definition of each criterion is not transparent enough and the evaluation procedures may be interpreted subjectively. In addition, the measurement scales of indicators and their aggregation mechanisms can be questionable, as well as for other models (Franceschini 2001; Franceschini et al. 2007).

5.4 Selection of a Synthetic Set of Indicators

The problem of selecting a synthetic set of indicators has been introduced in Sect. 4.6.2, when dealing with the properties of *exhaustiveness* and *non-redundancy*, and in Sect. 5.3, when presenting the performance dashboard, i.e., a tool that is supposed to synthetically represent a complex process through a relatively small number of indicators. Synthetic sets of indicators can be selected according to different operational approaches. Here we explore three of them:

1. Approach based on the concept of *relative importance*. This approach allows to select a (relatively small) set of indicators, focussing on the major representation targets.
2. Approach based on the so-called *minimum set covering*. This approach allows to select the smallest-possible set of indicators, providing an exhaustive representation of the process of interest.

Table 5.9 RADAR Scoring Matrix related to Enablers (EFQM 2013; APQI 2018). With permission

Element	Attribute score	0%	25%	50%	75%	100%
Approach	Sound: – approach with a clear rationale – well defined and well developed processes – approach focuses on stakeholder needs	No evidence	Low evidence	Evidence	Clear evidence	Complete evidence
	Integrated: – approach supports policy and strategy – when reasonable, approach is linked to other approaches	No evidence	Low evidence	Evidence	Clear evidence	Complete evidence
	Total	0 5 10	15 20 25 30 35	40 45 50 55 60	65 70 75 80 85	90 95 100
Deployment	Implemented: – approach is implemented	No evidence	Low evidence	Evidence	Clear evidence	Complete evidence
	Systematic: – approach implemented in a structured way	No evidence	Low evidence	Evidence	Clear evidence	Complete Evidence
	Total	0 5 10	15 20 25 30 35	40 45 50 55 60	65 70 75 80 85	90 95 100
Assessment and review	Measurement: – regular measurement of the effectiveness of the approach	No evidence	Low evidence	Evidence	Clear evidence	Complete evidence
	Learning: – periodical identification and sharing of good practices	No evidence	Low evidence	Evidence	Clear evidence	Complete evidence
	Improvement: – output information is used to identify, prioritise, plan and implement improvements	No evidence	Low evidence	Evidence	Clear evidence	Complete evidence
	Total	0 5 10	15 20 25 30 35	40 45 50 55 60	65 70 75 80 85	90 95 100
Total score		0 5 10	15 20 25 30 35	40 45 50 55 60	65 70 75 80 85	90 95 100

Table 5.10 RADAR Scoring Matrix related to Results (EFQM 2013; APQI 2018). With permission

Element	Attribute score	0%	25%	50%	75%	100%
Assessment and review	Trends: – Trends are positive and/or performances are generally good	No result	Positive trends and/or good performances for few results only	Positive trends and/or good performances for a certain number of results	Highly positive trends and/or excellent performances for most of the results	Highly positive trends and/or excellent performances in all relevant areas
	Targets: – Targets are achieved – Targets are appropriate	No result	Positive results in few areas only	Positive results in a certain number of areas	Positive results in most of the relevant areas	Excellent results in most of the relevant areas
	Comparisons: – Results compare well with acknowledged best-in-class organizations	No result	Comparisons in few areas only	Positive comparisons in a certain number of areas	Positive comparisons in most of the relevant areas	Excellent comparisons in most of the relevant areas
	Causes: – Results are a consequence of this approach	No result	Few results	Many results	Most of the results	All the results
	Scope: – Results address relevant areas	No result	In few areas	In many areas	In most of the areas	In all the areas
	Total	0 5 10	15 20 25 30 35	40 45 50 55 60	65 70 75 80 85	90 95 100

1. "Enablers" Criteria

Criterion number	1	%	2	%	3	%	4	%	5	%
Sub-criterion	1a		2a		3a		4a		5a	
Sub-criterion	1b		2b		3b		4b		5b	
Sub-criterion	1c		2c		3c		4c		5c	
Sub-criterion	1d		2d		3d		4d		5d	
Sub-criterion	1e		2e		3e		4e		5e	
Sum of percentage scores										
		÷ 5		÷ 5		÷ 5		÷ 5		÷ 5
Assigned score										

Note: assigned score is the arithmetic mean of sub-criteria percentage scores

2. "Results" Criteria

Criterion number	6			%	7			%	8			%	9			%
Sub-criterion	6a		x 0.75 =		7a		x 0.75 =		8a		x 0.25 =		9a		x 0.50 =	
Sub-criterion	6b		x 0.25 =		7b		x 0.25 =		8b		x 0.75 =		9b		x 0.50 =	
Assigned score																

3. Total score calculation

Criterion	Assigned score	%	Points
1 Leadership		x 10%	
2 Policy and strategy		x 10%	
3 People		x 10%	
4 Partnership and Resources		x 10%	
5 Processes		x 10%	
6 Customer results		x 14%	
7 People results		x 10%	
8 Society results		x 10%	
9 Key performance results		x 15%	
Total points			

Fig. 5.9 Scoring table related to RADAR Scoring Matrix (EFQM 2013; APQI 2018)

3. Approach based on the *degree of correlation*. The concept of degree of correlation is expressed in *qualitative* terms and indicates the influence of one indicator on one other (and *vice versa*) (Franceschini 2002). For example, the indicator 3.1 "Training of operators" – in Table 5.5 – may influence both indicators 3.3 "Customer perception of operator competence" and 6.1 "Response clarity". In fact, increasing training may encourage help-desk operators to be more competent and – consequently – to provide cleaner responses.

Establishing which of the above approaches is preferable depends on the peculiarities of (i) the process, (ii) the representation, and (iii) the data available.

5.4.1 The Concept of "Relative Importance"

The information contained in the Relationship Matrix—i.e., the tool introduced in Sect. 5.2.1, which links representation targets and indicators—can be used to determine a ranking of indicators, depending on their importance for the representation.

A classical method to obtain this ranking is the *Independent Scoring Method* (Akao 1988; Franceschini 2002), which includes two steps. The first one is the conversion of the symbolic relationships between representation targets and indicators into numerical values (Ø = 0, Δ = 1, O = 3, and ◉ = 9). This conversion is very delicate, as it hides a "promotion" from an ordinal scale (i.e., that of the relationship intensity) to a cardinal scale (see Sect. 3.2) (Franceschini and Romano 1999; Franceschini 2002).

The second step is the determination of the (absolute) importance of each (j-th) indicator, as:

$$w_j = \sum_{i=1}^{m} d_i \cdot r_{ij}, \tag{5.1}$$

where:

d_i is the degree of importance of the i-th representation target, ($i = 1, 2, \ldots, m$);
r_{ij} is the numerically-converted relationship between the i-th representation target and the j-th indicator;
w_j is the absolute importance of the j-th indicator ($j = 1, 2, \ldots, n$);
m is the number of representation targets;
n is the number of indicators.

The *absolute* importance (w_j) may be transformed into a *relative* importance, as:

$$w_j^* = \frac{w_j}{\sum_{j=1}^{n} w_j}. \tag{5.2}$$

Returning to the example presented in Sect. 5.2.1, symbolic (ordinal) relationships are firstly converted into numerical values. Then, the absolute/relative importance of indicators is calculated by applying the *Independent Scoring Method*; see the bottom of Fig. 5.10. For the purpose of example, the absolute importance of the first indicator is $w_j = 5{\cdot}9 + 4{\cdot}9 + 3{\cdot}1 + 5{\cdot}1 + 3{\cdot}4 = 101$, while the corresponding relative importance is $w_j^* = 101/588 = 17.18\%$.

A synthetic set of indicators can be selected considering those with the higher (relative) importance values. For example, considering the indicators with a relative importance of at least 10% ("cut treshold"), the following synthetic set is obtained:

- "Uniformity of responses" (21.09%);
- "Routing effectiveness" (17.18%);
- "Competence perception" (10.54%);
- "Confidential-data security" (10.20%);
- "Request-implementation time" (10.20%).

Key:
Ø=0 (no relationship)
Δ=1 (weak relationship)
O=3 (medium relationship)
●=9 (strong relationship)

		Indicators								
Represent. targets	Importance	Routing effectiveness	Accuracy in responses	Uniformity of responses (by different operators)	Request-implementation time	Competence perception	Percentage of answered calls	Courtesy of responses	Confidential-data security	Number of active lines
Reliability	5	●	O	●	O	Δ			O	
Responsiveness	5				●	Δ				
Competence	4	●	●	●		●				
Access	4						●			O
Courtesy	3			Δ				●		
Communication	3		Δ	●		Δ		O		
Credibility	3	Δ		O		O				
Security	5	Δ							●	
Understanding/ Knowing the customer	4	O	Δ	Δ		Δ		O		
Tangibles	3									●
Absolute importance		101	58	124	60	62	36	48	60	39
Relative importance (%)		17.18	9.86	21.09	10.20	10.54	6.12	8.16	10.20	6.63
Current values		93%	A	A	22 min	MB	98%	M	3%	3
Designed values		>90%	A	A	20 min	MB	>99%	A	<5%	5

Fig. 5.10 Calculation of the (absolute and relative) importance of indicators, for the Relationship Matrix related to a help-desk service. The intensity of relationships between representation targets and indicators is depicted by specific symbols: Ø (none), Δ, O, and ● (Franceschini 2002)

The "cut threshold" is a conventional value that may depend on the peculiarities of the process of interest or the expected number of indicators in the synthetic set. Of course, the higher the cut threshold, the lower the number of selected indicators.

Although the proposed approach allows to obtain a selection of the predominant indicators, it does not necessarily guarantee exhaustiveness and does not take into account the possible correlations of indicators.

The approaches presented in the next subsections will try to overcome these limitations.

5.4.2 The Concept of "Minimum Set Covering"

In some situations, it is required to define the *minimum set* of indicators that cover all the representation targets, providing a synthetic global vision of the process.

This is a classical combinatorial optimization problem, known as *set-covering problem* (Nemhauser and Wolsey 1988; Parker and Rardin 1988).

In more detail, given a set of elements {1, 2, ..., n} (called the universe) and a collection S of m sets, whose union equals the universe, the set-cover problem is to identify the smallest sub-collection of S, whose union equals the universe. For example, consider the universe $U = \{1, 2, 3, 4, 5\}$ and the collection of sets $S = \{\{1, 2, 3\}, \{2\}, \{3, 4\}, \{2, 4, 5\}\}$. Clearly, the union of S is U. However, we can cover all of the elements with the following smaller number of sets: {{1, 2, 3}, {2, 4, 5}}.

In formal terms, given a universe U and a family S of subsets of U, a *cover* is a subfamily $C \subseteq S$ of sets whose union is U. In the set-covering decision problem, the input is a pair (U, S) and an integer k, corresponding to the number of sets in S; the question is whether there is a set covering of size k or less. In the set-covering optimization problem, the input is a pair (U, S), and the task is to find a set covering that uses the fewest sets. The set-covering problem has a non-polynomial computational complexity, which increases with the problem dimension (Parker and Rardin 1988).

The search for the minimum number of indicators (e.g., those in the columns of the Relationship Matrix in Fig. 5.10) that are able to cover all representation targets (e.g., those in the rows of the Relationship Matrix in Fig. 5.10) can be interpreted as a particular *set-covering* problem. In the rest of the section, we present a heuristic algorithm, with polynomial complexity, which allows to quickly obtain a (sub-optimal) solution to this problem. This algorithm is known as Nemhauser's (Nemhauser and Wolsey 1988).

Nemhauser's Algorithm

This algorithm, which can be adapted to the problem of interest, can be summarized in four points:

1. Considering a Relationship Matrix, like the one exemplified in Fig. 5.10, select the indicator with the maximum number of relationships with representation targets (the relationship intensity is not considered, i.e., weak, medium, strong); in case of multiple indicators, the one with lowest *cost* can be selected (cf. notion of *economic impact* on Chap. 4).
2. The selected indicator is then removed from the Relationship Matrix, and included in the Critical Few set.
3. For the remaining indicators, the symbolic relationships related to the representation targets covered by the indicator selected at step 2 are removed.
4. The procedure is repeated until all the symbolic relationships in the Relationship Matrix are removed.

Returning to the *help-desk* example (Sect. 5.2.1), let us consider the Relationship Matrix in Fig. 5.11. In the first step, two possible indicators can be selected: "Uniformity of responses" and "Competence perception", since they both have six relationships with representation targets. For simplicity, assuming that the indicators' cost is the same, we select the first one and include it into the Critical Few set.

Then, all the symbols related to the representation targets covered by the selected indicator (i.e., representation targets 1, 3, 5, 6, 7, 9) are removed (see Fig. 5.12). The procedure is then reiterated considering the new Relationship Matrix. The next

		Indicators								
Represent. targets	Importance	Routing effectiveness	Accuracy in responses	Uniformity of responses (by different operators)	Request-implementation time	Competence perception	Percentage of answered calls	Courtesy of responses	Confidential-data security	Number of active lines
Reliability	5	◉	O	◉	O	Δ			O	
Responsiveness	5				◉	Δ				
Competence	4	◉	◉	◉		◉				
Access	4						◉			O
Courtesy	3			Δ				◉		
Communication	3		Δ	◉		Δ		O		
Credibility	3	Δ		O		O				
Security	5	Δ							◉	
Understanding/ Knowing the customer	4	O	Δ	Δ		Δ		O		
Tangibles	3									◉

Fig. 5.11 Application of Nemhauser's algorithm. The first step is the identification of the indicator with the largest number of relationships with representation targets. In this example, we assume that indicators have the same cost

selected indicator is "Number of active lines", since it has two (residual) relationships with representation targets. The new Relationship Matrix is shown in Fig. 5.13.

Among the four possible remaining indicators in the Relationship Matrix (i.e., "Routing effectiveness", "Request-implementation time", "Competence perception" and "Confidential-data security", see Fig. 5.13), we select the first one—"Routing effectiveness" – and we include it in the Critical Few set. Figure 5.14 shows the new Relationship Matrix.

Finally, between the two (residual) indicators ("Request-implementation time" and "Competence perception"), we select the first one.

In conclusion, the Critical Few indicators set is given by:

- "Uniformity of responses ";
- "Number of active lines";
- "Routing effectiveness";
- "Request-implementation time".

It can be noticed that Nemhauser's algorithm does not take into account neither the importance of representation targets nor the intensity of the relationhips (weak, medium, strong) between indicators and representation targets. These limitations are overcome by an enhanced version of this algorithm (Franceschini et al. 2007). However, this other version still ignores the (possible) correlations among indicators.

		Indicators								
Represent. targets	Importance	Routing effectiveness	Accuracy in responses	Uniformity of responses (by different operators)	Request-implementation time	Competence perception	Percentage of answered calls	Courtesy of responses	Confidential-data security	Number of active lines
Reliability	5									
Responsiveness	5				◉	Δ				
Competence	4									
Access	4						◉			O
Courtesy	3									
Communication	3									
Credibility	3									
Security	5	Δ							◉	
Understanding/ Knowing the customer	4									
Tangibles	3									◉

Fig. 5.12 Application of Nemhauser's algorithm. The second step is the identification of the indicator (among the remaining ones) that covers the highest number of representation targets (see the indicator in light grey). Dark-grey columns refer to the indicator(s) already included in the Critical Few set

5.4.3 The Concept of "Degree of Correlation"

Two indicators are correlated if variations in the first one determine variations in the second one and *vice versa*. (Potential) correlations among indicators are often determined on the basis of *qualitative* considerations. For example, when considering the so-far-exemplified Relationship Matrix (in Fig. 5.10), potentially correlated indicators tend to be related to the same representation targets (Franceschini 2002).

This idea can be developed to support the identification of correlations among indicators; the rationale is relatively simple: if an i-th indicator is related to a specific representation target, it will be likely to be correlated with another j-th indicator that is related to the same representation target. However, this is a *necessary* but *not sufficient* condition to state that two indicators are correlated. In other words, the fact that two indicators are related to similar representation targets is not necessarily a proof of their *real* correlation.

We now focus the attention on a (quantitative) semi-automatic procedure to support the search for potential correlations between indicators. We associate a column vector to each j-th indicator $\mathbf{b}_j \in \mathfrak{R}^n$ (n being the number of representation targets). Supposing that the Relationship Matrix $\mathbf{R}$ includes the symbols Δ, O and ◉ to depict weak, intermediate and strong relationships respectively, the coefficients of vectors $\mathbf{b}_j$ ($\forall j = 1, \ldots, m$) are determined as follows:

		Indicators								
Represent. targets	**Importance**	Routing effectiveness	Accuracy in responses	Uniformity of responses (by different operators)	Request-implementation time	Competence perception	Percentage of answered calls	Courtesy of responses	Confidential-data security	Number of active lines
Reliability	5									
Responsiveness	5				◉	Δ				
Competence	4									
Access	4									
Courtesy	3									
Communication	3									
Credibility	3									
Security	5	Δ							◉	
Understanding/ Knowing the customer	4									
Tangibles	3									

Fig. 5.13 Application of Nemhauser's algorithm. The third step is the identification of the indicator (among the remaining ones) that covers the highest number of representation targets (see indicators in light grey). Dark-grey columns refer to the indicators already included in the Critical Few set

$\forall i, j$ ($i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$)
if $r_{i,j} =$ ◉ (strong relationship) then $b_{i,j} = 9$;
if $r_{i,j} =$ O (intermediate relationship) then $b_{i,j} = 3$;
if $r_{i,j} = \Delta$ (weak relationship) then $b_{i,j} = 1$;
otherwise (absence of a relationship) $b_{i,j} = 0$.

Thus, a new binary matrix $\mathbf{B} \in \mathfrak{R}^{m,\ n}$ can be derived from matrix **R**. Columns vectors ($\mathbf{b}_j$) of Matrix **B** are then normalized producing a new matrix $\mathbf{N} \in \mathfrak{R}^{m,\ n}$, consisting of the column vectors $\mathbf{v}_j$ ($\forall j = 1, \ldots, m$), whose components are obtained as follows:

$$v_{i,j} = \frac{b_{i,j}}{\sqrt{\sum_{i=1}^{n} b_{i,j}^2}} \qquad \forall i, j. \tag{5.3}$$

Figure 5.15 exemplifies the construction of matrix **N**. The coefficient $q_{i,j}$ (i.e., scalar product of two generic column vectors $\mathbf{v}_i$ and $\mathbf{v}_j$) is introduced to represent the potential correlation between pairs of indicators:

Represent. targets	Importance	Routing effectiveness	Accuracy in responses	Uniformity of responses (by different operators)	Request-implementation time	Competence perception	Percentage of answered calls	Courtesy of responses	Confidential-data security	Number of active lines
		Indicators								
Reliability	5									
Responsiveness	5				◉	Δ				
Competence	4									
Access	4									
Courtesy	3									
Communication	3									
Credibility	3									
Security	5									
Understanding/ Knowing the customer	4									
Tangibles	3									

Fig. 5.14 Application of Nemhauser's algorithm. The fourth step is the identification of the indicator (among the remaining ones) that covers the highest number of representation targets (see indicators in light grey). Dark-grey columns refer to the indicators already included in the Critical Few set

$$R = \begin{array}{|c|c|c|}\hline \odot & \mathrm{O} & \\ \hline \Delta & & \mathrm{O} \\ \hline & \Delta & \\ \hline \end{array} \quad B = \begin{array}{|c|c|c|}\hline 9 & 3 & 0 \\ \hline 1 & 0 & 3 \\ \hline 0 & 1 & 0 \\ \hline \end{array} \quad N = \begin{array}{|c|c|c|}\hline 0.99 & 0.95 & 0.00 \\ \hline 0.11 & 0.00 & 1.00 \\ \hline 0.00 & 0.32 & 0.00 \\ \hline \end{array}$$

Fig. 5.15 Example of construction of matrix **N**; symbols in the corresponding matrix **R** are converted according to the standard numerical conversion: Ø = 0, Δ = 1, O = 3, and ◉ = 9

$$q_{i,j} = q_{j,i} = \mathbf{v}_i \cdot \mathbf{v}_j \qquad \forall i,j = 1, \ 2, \ \ldots, \ m. \tag{5.4}$$

By calculating $q_{i,j}$ for all the possible pairs of indicators (and relevant column vectors in the **N** matrix), it is possible to determine the correlation matrix **Q**:

$$\mathbf{Q} = \mathbf{N}^T\mathbf{N} \tag{5.5}$$

$\mathbf{Q} \in \Re^{m,m}$ is square and symmetrical, with $q_{i,\ i} = 1, \forall i = 1, \ 2, \ldots, \ m$.

Table 5.11 Extract of the Relationship Matrix (**R**) in Fig. 5.10

R =	⊙	O	⊙	O	Δ			O	
				⊙	Δ				
	⊙	⊙	⊙		⊙				
						⊙			O
			Δ				⊙		
		Δ	⊙		Δ		O		
	Δ		O		O				
	Δ							⊙	
	O	Δ	Δ		Δ		O		
									⊙

Table 5.12 Matrix **B**, containing the numerical values related to the symbols in Table 5.11, according to the standard conversion Ø = 0, Δ = 1, O = 3, and ⊙ = 9

B =	9	3	9	3	1			3	
				9	1				
	9	9	9		9				
						9			3
			1				9		
		1	9		1		3		
	1		3		3				
	1							9	
	3	1	1		1		3		
									9

Matrix **Q** expresses the degree of correlation of indicators, in terms of their ability to impact on the same representation targets.

The $q_{i,j}$ values (contained in **Q**) are compared with a conventional threshold t (with $0 \leq t \leq 1$); $\forall i, j$, if $q_{i,j} > t$ then a *potential* correlation between the i-th and j-th indicator is revealed. The resulting information about potential correlations is included into a new matrix $\widehat{\mathbf{Q}}$, including the symbol "X" for potential correlations.

Considering the call-center example (Fig. 5.10), we obtain the matrices **R**, **B**, **N** and **Q**, which are respectively shown in Tables 5.11, 5.12, 5.13 and 5.14. Setting the threshold value to $t = 0.75$, we obtain the correlation matrix $\widehat{\mathbf{Q}}$ in Fig. 5.16.

Starting from matrix $\widehat{\mathbf{Q}}$, a set of Critical Few indicators can be selected according to the following four-step procedure:

1. Include those indicators that are not related to other indicators into the Critical Few set; then remove these indicators from the $\widehat{\mathbf{Q}}$ matrix (i.e., delete the elements in the corresponding rows and columns).
2. Among the remaining indicators, select the one with the highest number of correlations with other ones; in case of multiple indicators, select the one of lowest cost.

3. Include the indicator selected at step (2) into the Critical Few set; then, remove this indicator and those correlated with it from matrix $\widehat{\mathbf{Q}}$ (i.e., delete the elements in the corresponding rows and columns).
4. Repeat the procedure starting from step (2), till all indicators have been removed from matrix $\widehat{\mathbf{Q}}$.

The application of the procedure to the correlations in Fig. 5.16 (t is set to 0.75) produces the following Critical Few set (in one iteration only):

- "Request-implementation time", "Percentage of answered calls", "Courtesy of responses", "Confidential-data security" and "Number of active lines", which are selected at step (1), being not correlated with each other.
- "Routing effectiveness", since it is correlated with four other indicators (see Fig. 5.16) and reasonably less expensive than the "Accuracy in responses" indicator (correlated with four other indictors too).

The afore-described procedure does not guarantee a complete coverage of representation targets by the selected indicators, i.e., each representation target will not necessarily have (at least) a relationship with (at least) an indicator of the Critical Few set (e.g., consider the verification procedure in Fig. 5.10). In general, the lower the (conventional) value of t, the larger the number of (potential) correlations among indicators, and the lower the number of indicators in the Critical Few set, with a consequent reduction in the probability of guaranteeing complete coverage.

This problem can be overcome by introducing some adaptations: the procedure can be iteratively applied, imposing a relatively low initial t value and—if necessary—increasing it, until a complete covering of representation targets is achieved.

Following this idea, the procedure can be enhanced as follows:

1. Set a relatively low value of t (e.g., 0.2).
2. Construct matrix $\widehat{\mathbf{Q}}$ according to the t value set at step (1).
3. Include those indicators that are not related to other indicators into the Critical Few set; then remove these indicators from the $\widehat{\mathbf{Q}}$ matrix (i.e., delete the elements in the corresponding rows and columns).
4. Among the remaining indicators, select the one with the highest number of correlations with other ones; in case of multiple indicators, select the one of lowest cost.
5. Include the indicator selected at step (4) into the Critical Few set; then, remove this indicator and those correlated with it from matrix $\widehat{\mathbf{Q}}$ (i.e., delete the elements in the corresponding rows and columns).
6. Repeat the procedure starting from step (4), till all indicators have been removed from matrix $\widehat{\mathbf{Q}}$.
7. Check whether the selected indicators guarantee a complete covering of representation targets (e.g., using the procedure seen before). If so, the procedure

Table 5.13 Matrix **N**, aggregating the normalized column vectors from matrix **B** (see Table 5.12)

N =	0.68	0.31	0.56	0.32	0.10	0.00	0.00	0.32	0.00
	0.00	0.00	0.00	0.95	0.10	0.00	0.00	0.00	0.00
	0.68	0.94	0.56	0.00	0.93	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.32
	0.00	0.00	0.06	0.00	0.00	0.00	0.90	0.00	0.00
	0.00	0.10	0.56	0.00	0.10	0.00	0.30	0.00	0.00
	0.08	0.00	0.19	0.00	0.31	0.00	0.00	0.00	0.00
	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.00
	0.23	0.10	0.06	0.00	0.10	0.00	0.30	0.00	0.00
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95

Table 5.14 Matrix (**Q**) containing the correlation coefficients, obtained from matrix **N** (Table 5.13)

Q =	1.00	0.88	0.80	0.22	0.75	0.00	0.07	0.29	0.00
	0.88	1.00	0.77	0.10	0.92	0.00	0.06	0.10	0.00
	0.80	0.77	1.00	0.18	0.71	0.00	0.25	0.18	0.00
	0.22	0.10	0.18	1.00	0.13	0.00	0.00	0.10	0.00
	0.75	0.92	0.71	0.13	1.00	0.00	0.06	0.03	0.00
	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.32
	0.07	0.06	0.25	0.00	0.06	0.00	1.00	0.00	0.00
	0.29	0.10	0.18	0.10	0.03	0.00	0.00	1.00	0.00
	0.00	0.00	0.00	0.00	0.00	0.32	0.00	0.00	1.00

	(I_1) Routing effectiveness	(I_2) Accuracy in responses	(I_3) Uniformity of responses	(I_4) Request-implementation time	(I_5) Competence perception	(I_6) Percentage of answered calls	(I_7) Courtesy of responses	(I_8) Confidential-data security	(I_9) Number of active lines
(I_1) Routing effectiveness	X	X	X		X				
(I_2) Accuracy in responses	X	X	X		X				
(I_3) Uniformity of responses	X	X	X						
(I_4) Request-implementation time				X					
(I_5) Competence perception	X	X			X				
(I_6) Percentage of answered calls						X			
(I_7) Courtesy of responses							X		
(I_8) Confidential-data security								X	
(I_9) Number of active lines									X

Fig. 5.16 Matrix $\widehat{\mathbf{Q}}$ identifying the (potential) correlations among the indicators in Fig. 5.10. In this specific case, t is set to 0.75. Correlations are depicted by the symbol "X"

stops; otherwise threshold t is increased a little (e.g., by imposing an increase of 0.1 or 0.2) and the procedure restarts from point (2).

An application example of the enhanced procedure is illustrated in (Franceschini et al. 2007). The scientific literature includes other heuristic methodologies, which can be used to select Critical Few indicators.

5.5 Implementing a System of Performance Indicators

This section introduces some guidelines to develop a system of performance indicators for a generic organization. Description is supported by several practical examples.

Step 1 *Establishing the working group that will activate the performance measurement system.*

Step 2 *Defining a proper terminology within the organization.*

Apart from the indicator classification in Sects. 1.3 and 3.5, indicators can be divided into five categories:

- *Input indicators*: are used to understand the human and capital resources needed to produce outputs and outcomes.
- *Process indicators*: are used to understand the intermediate steps in producing a product/service.
- *Output indicators*: are used to measure the product/service provided by the system or organization and delivered to customers/users.
- *Outcome indicators*: evaluate the expected, desired, or actual result(s) to which the outputs of the activities of a service or organization have an intended effect.
- *Impact indicators*: measure the direct or indirect effects or consequences resulting from achieving program goals.

A second possible classification is based on the moment in which indicators are constructed. These types of indicators are defined below:

- *Lagging indicators*: measure performance after the fact. Project cost performance is an example of a lagging indicator used to measure program performance.
- *Leading indicators*: are more predictive of future performance. They include, for example, estimated costs.
- *Behavioural indicators*: measure the underlying culture or attitude of the personnel or organization being measured. An example is given by questionnaires concerning the satisfaction of employees.

Step 3 *Design general criteria.*

Here are some general criteria to take into account when developing a performance measurement system:

- Keep the number of performance indicators to a minimum. For any program, there is a large number of potential performance indicators. It is important to identify a limited number of "critical" indicators;
- Process objectives must be understandable and must be developed clearly. Experience has shown that performance measurement systems frequently fail because the respective parties do not have a common understanding of the meaning and goals of indicators;
- Determine if the cost of the performance indicator is worth the gain. Sometimes the cost of obtaining an indicator may outweigh any added value resulting from the use of the indicator itself (cf. concept of *economic impact* in Sect. 4.6.1);
- Assure that indicators are comprehensive. In developing performance indicators, consider measuring positive performance as well as minimizing possible negative side-effects of the program (cf. concept of *simplicity of use* in Sect. 4.6.1);
- Consider a risk evaluation. Organizations should consider a risk evaluation of the organization to determine which specific processes are most critical to organizational success or which processes pose the greatest risk to successful mission accomplishments;
- Place greater emphasis on measuring the risk produced by the use of a particular performance indicator, both for short and long term;
- Consider the weight of conflicting performance indicators. For example, an objective of high productivity may conflict with an objective for a high quality product (cf. concept of *counter-productivity* in Sect. 4.6.1);
- Develop consistent performance indicators that promote teamwork. Performance indicators should be designed to maximize teamwork. The performance indicators for different levels of an organization should be generally consistent with each other, from top to bottom and across the hierarchy. The risk of sub-optimization should be determined when setting performance indicators.

Step 4 *How to check performance indicators.*

After having developed a system of performance indicators, it is important to check/test it. Here are several possible ways of doing this.

SMART test (University of California 2017)

- **S** (*Specific*): is the indicator clear and focused, so it avoids misinterpretation? It should include measurement assumptions and definitions, and should be easily interpreted.
- **M** (*Measurable*): can the indicator be quantified and compared to other data? It should allow for meaningful statistical analysis.

- **A** (*Attainable*): is the indicator achievable, reasonable and credible under expected conditions?
- **R** (*Realistic*): does the indicator fit into the organization's constraints? Is it cost-effective?
- **T** (*Timely*): can the indicator be evaluated within the given time frame?

The "*Three Criteria*" test (U.S. Department of Energy – PBM SIG 2012)

Another test of performance indicators includes the verification of three general criteria:

- *Strategic Criteria*—do the indicators enable strategic planning and then drive the deployment of the actions required to achieve objectives and strategies? Do the indicators align behaviour and initiatives with strategy and focus the organization on its priorities?
- *Quantitative Criteria*—do the indicators provide a clear understanding of progress toward objectives and strategy as well as the current status, rate of improvement, and probability of achievement? Do the indicators identify gaps between current status and performance aspirations, thereby highlighting improvement opportunities?
- *Qualitative Criteria*—are the indicators perceived as valuable by the organization and people involved?

The Treasury Department Criteria test (U.S. Department of the Treasury 1994)

The test is based on the following general verification criteria:

1. **Data criteria**—data availability and reliability can affect the selection and development of performance indicators.
   - *Availability*: are data required by the indicators in use available? If not, can these data be collected? Is there any indicator that is better than the ones in use, according to the data available?
   - *Accuracy*: are data sufficiently reliable? Is there any bias, exaggeration, omission or error that is likely to make (some of the) indicators inaccurate or misleading?
   - *Timeliness*: do data available allow us to obtain a responsive evaluation of the process performance? How frequently should we collect and/or report data (e.g., monthly vs. annually)?
   - *Security*: is there any privacy concern preventing the use of these data by third parties?
   - *Costs of data collection*: are the available resources (i.e., know how, computer capability, fundings, etc.) appropriate to data collection? Is data collection cost-effective?
2. **Indicator criteria**
   - *Validity*: does the indicator address financial or program results? Can changes in the value of the indicator be clearly interpreted as desirable or undesirable?

Does the indicator clearly reflect changes in the program? Is there a sound, logical relationship between the program and what is being measured, or are there significant uncontrollable factors?
- *Uniqueness*: does the information conveyed by one indicator duplicate information provided by another (cf. concept of *non-redundancy* in Sect. 4.6.2)?
- *Evaluation*: are there reliable benchmark data, standards, or alternative references for interpreting the selected performance indicators?

3. **Measurement system criteria**
   - *Balance*: is there a balance between input, output, or outcome indicators, and productivity or cost-effectiveness indicators? Does the mix of indicators compensate for the bias in any single indicator (cf. property of *compensation* in Sect. 4.6.3)?
   - *Completeness*: are the major process dimensions covered? Do indicators cover the major objectives (cf. property of *exhaustiveness* in Sect. 4.6.2).
   - *Usefulness*: does the organization management use the proposed indicators? Is there any incentive for management to use indicators? Is the management trained to use and interpret the indicator results?

It is worth remarking that some of the afore-described criteria are equivalent to some of the properties discussed in Chap. 4.

Step 5 *Benchmarking with the performance measurement systems of other organizations*.

The point here is relatively simple: avoiding to "reinvent the wheel", i.e., avoiding repeating the errors made by other organizations.

5.5.1 Examples of Developing Performance Indicators

Many different approaches for developing performance measurement systems have been proposed in the scientific literature. Two of them are presented in the following sections:

- The Auditor General of Canada approach;
- The DOE/NV approach (U.S. Department of Energy/Nevada Operations 1994).

The Auditor General of Canada Approach

This approach is illustrated in detail in the document "Developing Performance Measures for Sustainable Development Strategies", produced by the Auditor General of Canada and the Commissioner of the Environment and Sustainable Development (2017). This approach is supposed to assist work units in developing objectives and indicators that contribute to achieving the strategic objectives for sustainable development of their department. This approach relies on the idea that there is a direct link between the strategic objectives of a department and the specific activities of the relevant work units.

The methodology includes two main parts:

1. Identifying the objectives of work units, which may contribute to strategic objectives. This activity contains five work steps (1–5).
2. Establishing performance indicators. This activity includes additional four steps (6–9).

Part 1

A performance framework guides performance planning, identifying the possible links between activities, outputs and results. A good performance framework should address the following questions:

- *Why* is the program of the work unit relevant to strategic objectives? Special attention should be given to the implications for long term and sustainable development.
- *Who* are the subjects involved? E.g., target groups, stakeholders, etc.
- *What* result do we expect to achieved? Special attention should be given to the impact of short/medium term activities on the strategic objectives of the department.
- *How* can strategic objectives be achieve? This question concerns program inputs, processes, activities and outputs.

Step 1 *Defining the program*

Defining the role of the program for strategic objectives is essential for the clear definition of targets and performance indicators. For the purpose of example, Table 5.15 links program activities to strategic objectives for a program copncerning the reduction of defectives in a production plant of exhaust systems (see Sect. 3.6.1).

Step 2 *Identifying key program activities and outputs*

This step is essential to identify key issues for achieving strategic objectives. Table 5.16 schematizes a possible procedure; relationships among activities/outputs and strategic objectives are defined using a three-level qualitative scale: High (H), Medium (M) or Low (L). The last column synthesizes the overall impact of activities/outputs, through the following operator (Franceschini et al. 2005):

$$L_{AGG} = \min_{o_i \in O} \{L(o_i)\} \tag{5.6}$$

being $L(o_i)$ the impact of the activity/output of interest on the strategic objective o_i, and O the whole set of strategic objectives.

Table 5.15 Example of link between program activities and strategic objectives, for a program concerning the reduction of defectives in a production plant of exhaust systems (see Sect. 3.6.1)

Program activities	Description	Strategic objectives or outcomes
Activity 1: Personnel training	Increasing competence, skill and participation of personnel	Reduction of human errors
Activity 2: Planning preventive maintenance	Increasing reliability of manufacturing processes	Reduction of the causes of process failures
Activity 3: Introduction of a control system	Facilitating the quick detection of process failures	Process control

Table 5.16 Table linking program activities/outputs and strategic objectives. The relevant relationships may have three intensity levels: High (H), Medium (M) or Low (L). Last column includes the aggregated importance of each activity/output. Activities/outputs and strategic objectives are defined in Table 5.15

Program activities and outputs	Strategic objectives			Aggregated importance L_{AGG}
	Reduction of human errors	Reduction of the causes of process failures	Process control	
Activity 1: Personnel training	H	H	H	H
Output 1: Increasing personnel's competence	H	M	M	M
Activity 2: Planning preventive maintenance	L	H	L	L
Output 2: Reducing the number of process failures	L	H	M	L
Activity 3: Introduction of a control system	H	H	H	H
Output 3: Reducing time to detect process failures	M	M	H	M

Step 3 *Identify program stakeholdersand issues*

Before focussing on strategic objectives, we should identify: users and other stakeholders, who are in some way affected by the program activities/outputs. For significant program activities/outputs, Table 5.17 identifies their link to stakeholders and issues.

Step 4 *Identifying the expected results of the program*

Table 5.18 establishes a connection between activities/outputs and expected (medium/long-term) results.

Table 5.17 Identifying key issues and stakeholders related to the activities/outputs in Table 5.16. Activities and outputs are listed in order of importance for strategic objectives (High, Medium, Low)

Program activities and outputs	Key issues		Stakeholders	
	Desired program effects	Undesired program effects	Positively affected	Negatively affected
Activity 1 (H): Personnel training	Reducing defectiveness due to human errors; encouraging personnel commitment	Additional cost for training personnel	Process operators	Production director
Activity 3 (H): Introduction of a control system	Detecting the causes of potential failures	Increasing process time; need for dedicated staff	Process operators, process leaders	Production director
Output 1 (M): Increasing personnel's competence	Increasing the personnel's competence and skill	Creating expectation of career advancement	Process operators	None
Output 3 (M): Reducing time to detect process failures	Quick solution to problems	None	Process operators, process leaders	None
Activity 2 (L): Planning preventive maintenance	Increasing process reliability	Additional cost for analysing process reliability and performing maintenance operations	Process leaders	None
Output 2 (L): Reducing the number of process failures	Increasing process reliability	None	Process leaders	None

Step 5 *Identifying performance requirements*

Performance requirements must be defined in operational terms in order to be managed effectively. Table 5.19 establishes a connection between performance requirements and expected results (operationalization of objectives).

Part 2

The next four steps drive the definition of sound performance indicators.

Step 6 *Identifying potential performance indicators*

Table 5.18 Defining results related to the activities/outputs in Table 5.17. Activities and outputs are listed in order of importance for strategic objectives (High, Medium, Low)

	Expected results (objectives)	
Program activities and outputs	Medium-term	Long-term
Activity 1 (H): Personnel training	Better knowledge of process issues	Highly qualified personnel
Activity 3 (H): Introduction of a control system	Establishing a process control system	Process completely under control
Output 1 (M): Increasing personnel's competence	(At least) 50% of the operators are qualified	All operators are qualified
Output 3 (M): Reducing time to detect process failures	Establishing a failure-detection system	Enhancing the failure-detection system
Activity 2 (L): Planning preventive maintenance	Analysing process reliability	Efficient preventive-maintenance system
Output 2 (L): Reducing the number of process failures	Significant reduction of process failures	Total absence of process failures

Table 5.19 Defining performance requirements related to the expected results (from Table 5.18). The analysis is limited to long-term results but it can be extended to the medium-term ones

(Long-term) expected results	Program activities and outputs	Performance requirements
Highly qualified personnel	*Activity 1* (H): Personnel training	Organization of training activities (selecting courses, recruiting teachers, etc.)
Process completely under control	*Activity 3* (H): Introduction of a control system	Definition of control procedures, tools (e.g., control charts and relevant parameters)
All the operators are qualified	*Output 1* (M): Increasing personnel's competence	Construction of a system to evaluate personnel's competence
Enhancing the failure-detection system	*Output 3* (M): Reducing time to detect process failures	Construction of a system for data collection and analysis
Efficient preventive-maintenance system	*Activity 2* (L): Planning preventive maintenance	Construction of a detailed maintenance plan
Total absence of process failures	*Output 2* (L): Reducing the number of process failures	Construction of a system to analyse process reliability

Performance indicators should allow to identify the gap between actual performance and target performance.

Performance indicators represent an important feedback for management. Table 5.20 links expected results, performance requirements, and (possible) indicators.

Table 5.20 Identifying (potential) performance indicators related to the performance requirements defined in Table 5.19. The analysis is limited to long-term results but it can be extended to the medium-term ones

(Long-term) expected results	Activities and outputs	Performance requirements	(Potential) performance indicators
Highly qualified personnel	*Activity 1* (H): Personnel training	Organization of training activities (selecting courses, recruiting teachers, etc.)	Percentage of highly qualified employees, for each process activity
Process completely under control	*Activity 3* (H): Introduction of a control system	Definition of control procedures, tools (e.g., control charts and relevant parameters)	Estimated percentage of unnoticed defects (in a specific time window)
All operators are qualified	*Output 1* (M): Increasing personnel's competence	Construction of a system to evaluate personnel's competence	Percentage of "very competent" employees
Enhancing the failure-detection system	*Output 3* (M): Reducing time to detect process failures	Construction of a system for data collection and analysis	Time between failure occurrence and detection
Efficient preventive-maintenance system	*Activity 2* (L): Planning preventive maintenance	Construction of a detailed maintenance plan	Number of process failures (in a monthly period) due to lack of programmed maintenance
Total absence of process failures	*Output 2* (L): Reducing the number of process failures	Construction of a system to analyse process reliability	Total number of process failures (in a monthly period)

Step 7 *Understanding information capabilities related to indicators*

Before selecting performance indicators, organizations should consider their data-collection and data-analysis capabilities. Table 5.21 can be used to support this activity.

Step 8 *Assessing the plausibility of performance indicators*

Once performance indicators are developed, we should check their plausibility and consistency with objectives.

Below is a selection of criteria to check plausibility of indicators:

- *Meaningful*
  - *understandable* (clearly and consistently defined, well explained, measurable, with no ambiguity);
  - *relevant* (i.e., consistent with objectives, practical to users, applicable to the activities of interest);

Table 5.21 Data-collection scheme. Performance indicators are those defined in Table 5.20

Potential performance indicators	Units	Initial value
Percentage of highly qualified employees, for each process activity	%	50
Estimated percentage of unnoticed defects (in a specific time window)	%	0
Percentage of "very competent" employees	%	50
Time between failure occurrence and detection	hours	24
Number of process failures (in a monthly period) due to lack of programmed maintenance	$\frac{\text{number}}{\text{month}}$	0
Total number of process failures (in a monthly period)	$\frac{\text{number}}{\text{month}}$	5

  - *comparable* (allows comparisons with other organizations/activities/standards).
- *Reliable*
  - represents what is supposed to be measured;
  - data required can be replicated;
  - data and analysis are free from error;
  - not easy to manipulate;
  - well balanced with respect to other indicators.
- *Practical*
  - Financially sustainable;
  - Responsive.

It is interesting to compare these requirements with those included in the taxonomy in Chap. 4.

Table 5.22 helps to select suitable indicators.

Step 9 *Accountability related to performance indicators*

Accountability means formalizing the relationships among results, outputs, activities, and resources. It allows employees to see how their activities may contribute to the success of the organization. Tables 5.23 and 5.24 provide a reference scheme which may support the accountability analysis.

According to the Process Auditor method, a set of performance indicators should provide a picture of results for managers, executives and internal/external stakeholders. Additionally, it explains how the committed resources contribute to achieve specific results; in this sense, it has a strong "constitutive" connotation.

The DOE/NV Approach

This section presents the methodology for establishing a performance measurement system proposed by DOE/NV (U.S. Department of Energy 1996a, b; U.S. Department of Energy – PBM SIG 2012). This methodology is applied to a practical case study as follows.

Table 5.22 Scheme for supporting the selection of suitable indicators. Performance indicators are those reported in Table 5.20

Performance indicators	Meaningful			Reliable	Practical
	Understandable	Relevant	Comparable		
Percentage of highly qualified employees, for each process activity	No	No	No	No	Yes
Estimated percentage of unnoticed defects (in a specific time window)	Yes	Yes	Yes	No	Yes
Percentage of "very competent" employees	No	Yes	No	Yes	Yes
Time between failure occurrence and detection	Yes	No	No	Yes	Yes
Number of process failures (in a monthly period) due to lack of programmed maintenance	No	Yes	No	No	No
Total number of process failures (in a monthly period)	Yes	Yes	Yes	Yes	Yes

The Communications & Information Management Company provides communications and information management services. The company's warehouse, part of the Property Management Division, provides storage and excess services for company property in the custody of 25 divisions. The warehouse department has a staff of ten people: a warehouse supervisor, four property specialists, one property clerk, three drivers, and one data-entry clerk. The warehouse makes approximately 50 pickups per week at company locations that include remote areas.

To request services from the warehouse, a division customer contacts the warehouse-property clerk, requesting a pickup of goods for storage or excess. The customer provides the clerk with the identification number or serial number for each good to be picked up and brought to the warehouse. There are typically 1–20 goods per pickup. If a pickup date is not requested by the customer, a date will be provided to the customer by the property clerk. The property clerk completes a property transfer form, which reflects the date of the call, customer's name, division, location, property identification number and date scheduled for pickup.

A goal of the warehouse is not to exceed 3 days from the date of the call to the time of the pickup, unless a special date has been requested by the customer. The warehouse receives approximately ten calls per week for pickups on special dates. On the scheduled pickup day, the assigned driver takes the transfer form to the designated location. The driver is responsible for ensuring that each good matches the property identification number or serial number listed on the transfer form. After the truck is loaded, the driver obtains the customer's signature on the transfer form.

Table 5.23 Establishing accountability related to the performance measurement system defined in Tables 5.19 and 5.21

(Long-terms) expected results	Responsible party for expected results	Activities and outputs	Responsible party for activities and outputs	Performance indicators	Responsible party for indicators
Highly qualified personnel	Personnel manager	*Activity 1* (H): Personnel training	Personnel manager (training area)	Percentage of highly qualified employees, for each process activity	Statisticians (personnel area)
Process completely under control	Production manager	*Activity 3* (H): Introduction of a control system	Quality manager	Estimated percentage of unnoticed defects (in a specific time window)	Process-control operators
All the operators are qualified	Personnel manager	*Output 1* (M): Increasing personnel's competence	Personnel manager (training area)	Percentage of "very competent" employees	Statisticians (personnel area)
Enhancing the failure-detection system	Production manager	*Output 3* (M): Reducing time to detect process failures	Quality manager	Time between failure occurrence and detection	Process-control operators
Efficient preventive-maintenance system	Production manager	*Activity 2* (L): Planning preventive maintenance	Maintenance manager	Number of process failures (in a monthly period) due to lack of programmed maintenance	Maintenance operators
Total absence of process failures	Production manager	*Output 2* (L): Reducing the number of process failures	Quality manager	Total number of process failures (in a monthly period)	Maintenance operators

Table 5.24 Identifying resources required for implementing the proposed performance measurement system (see Tables 5.19 and 5.20)

(Long-term) expected results	Program activities/outputs	Required resources		
		Human	Financial	Other
Highly qualified personnel	*Activity 1* (H): Personnel training	2	50,000 € per year	
Process completely under control	*Activity 3* (H): Introduction of a control system	2	65,000 € per year	
All operators are qualified	*Output 1* (M): Increasing personnel's competence			
Enhancing the failure-detection system	*Output 3* (M): Reducing time to detect process failures			
Efficient preventive-maintenance system	*Activity 2* (L): Planning preventive maintenance	3	100,000 € per year	
Total absence of process failures	*Output 2* (L): Reducing the number of process failures			

The driver also signs the form and provides the customer with a copy acknowledging the receipt.

The driver returns to the warehouse, where a property specialist annotates the date on the transfer form, unloads the truck, and provides the data-entry clerk with the signed copies of the form. The data-entry clerk enters the information from the transfer form into the automated accountable property system and the transfer forms are then filed. The data entered are intended to transfer accountability from the division customer to the warehouse. At the end of the month, division customers receive a computer-generated property list indicating the accountable property in their location for which they are responsible. The customer reviews this report for accuracy. If customer records contain some inaccuracies concerning a customer call, the warehouse supervisor logs a complaint including: date of the call, division name, property location, date of the property list, and description of discrepancies. The supervisor assigns a property specialist to resolve these discrepancies.

The group is responsible for many processes, such as delivering property, conducting inventory, etc. For the purpose of simplicity, the following description summarizes the operational steps to develop indicators for the process "goods pickup and storage". The work team involves the entire staff.

Step 1 *Process identification*

The first step consists in identifying process inputs, outputs, activities, and resources (see Sect. 5.2.2). Figure 5.17 provides a flow-chart representation of the process of interest. This activity includes the identification of the process objectives and outputs (see Fig. 5.18).

Process objectives are:

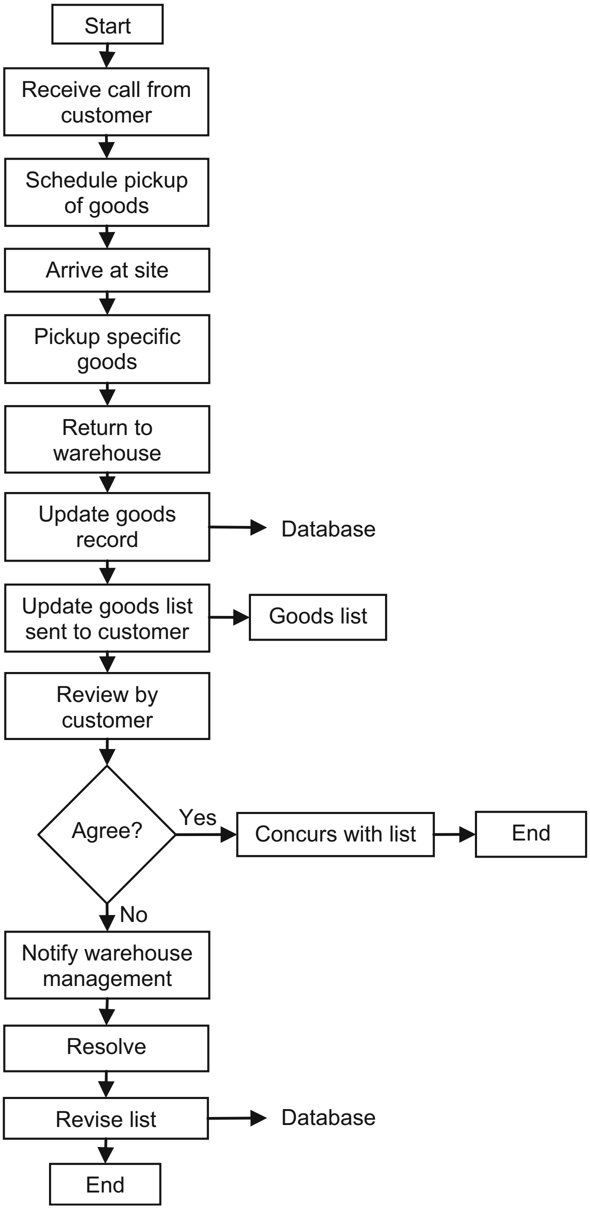

Fig. 5.17 Flow chart for the process of "pickup and storage of goods" (U.S. Department of Energy – PBM SIG 2012). With permission

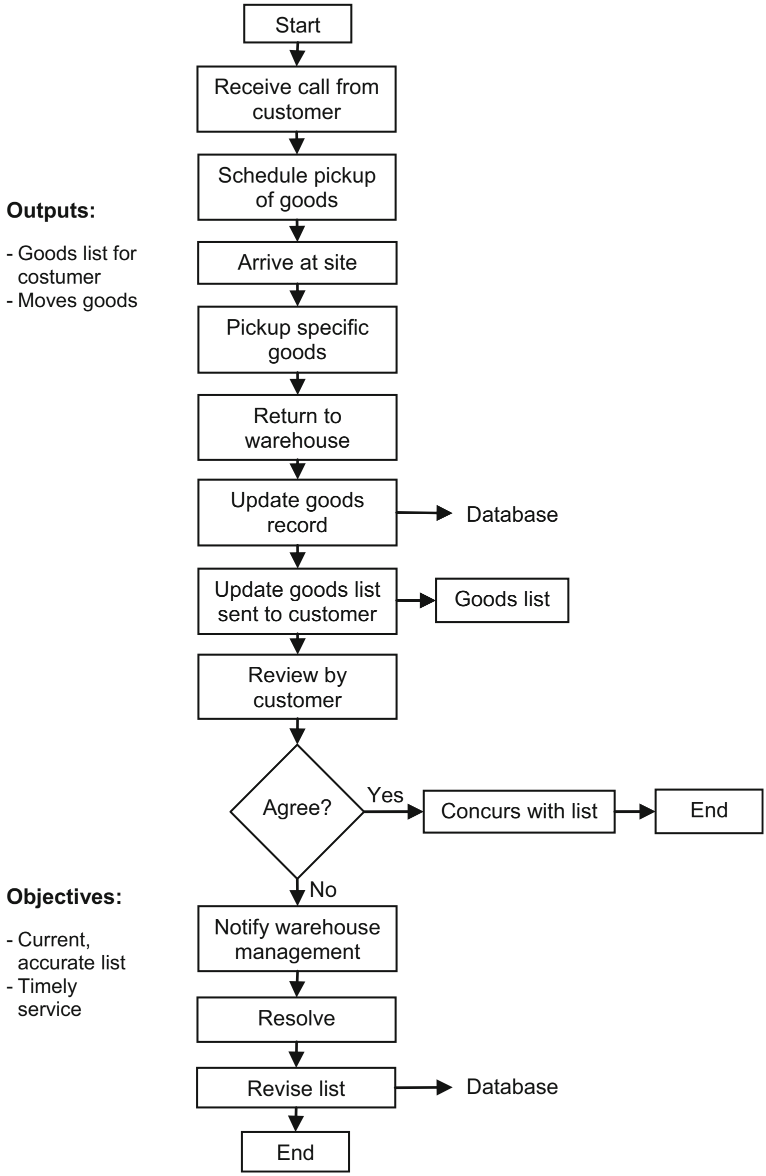

Fig. 5.18 Flow chart for the process of "pickup and storage of goods", supplied with the process objectives and outputs (U.S. Department of Energy – PBM SIG 2012). With permission

- a current, accurate goods list for customers;
- timely pickup and removal of goods.

Outputs are:

- a list of goods for customers;
- removal and storage of company goods.

Step 2 *Identification of critical activities*

The next step is to determine how objectives will be met. In this case, two critical activities, i.e., activities with lower-than-expected performance, have been identified; Consequently, two corresponding control points have been defined (see Fig. 5.19).

Step 3 *Establishing performance goals or standards*

For each control point, it is necessary to establish a performance goal or standard. For critical activity 1 ("*Return to warehouse*"), three goals have been defined (see Fig. 5.20):

- 3-day turnaround;
- Fulfill 100% of requests;
- 95% of requests fulfilled in time.

For critical activity 2 ("*Resolve discrepancies*") (see Fig. 5.20):

- 98% of the records in the monthly goods list should be accurate;
- No more than 5% of the time should be spent resolving discrepancies.

Step 4 *Establish performance indicators*

Performance indicators are supposed to represent important aspects of the process. We now identify specific performance indicators for the two critical activities. In particular, for **critical activity 1** ("*Return to warehouse*"), we define the following indicators (see Fig. 5.21):

- *Performance indicator 1-A*: "Number of days between (pickup) request and actual pickup".
  - Collected data: pickup-request date and (actual) pickup date.
  - Means of collection: goods transfer form.
  - Frequency: weekly.
- *Performance indicator 1-B*: "Percentage of special pickups fulfilled in time":

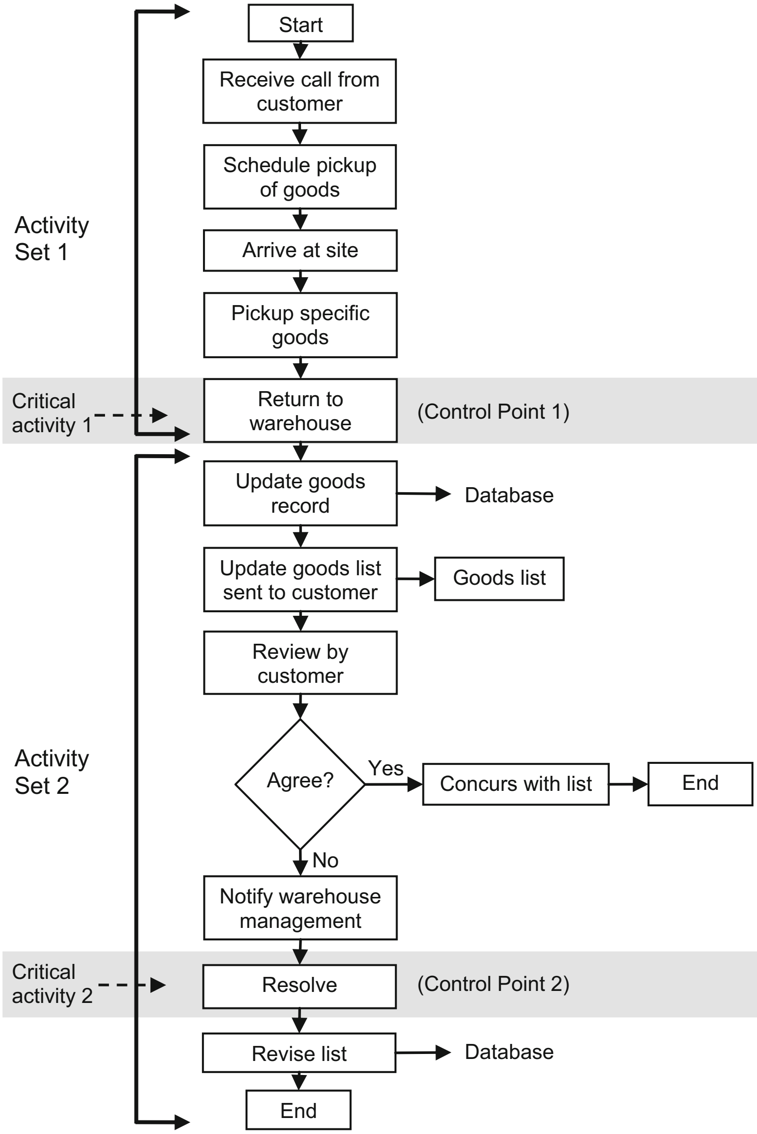

Fig. 5.19 Flow chart for the process of "goods pickup and storage", supplied with the representation of the process critical activities and control points (U.S. Department of Energy – PBM SIG 2012). With permission

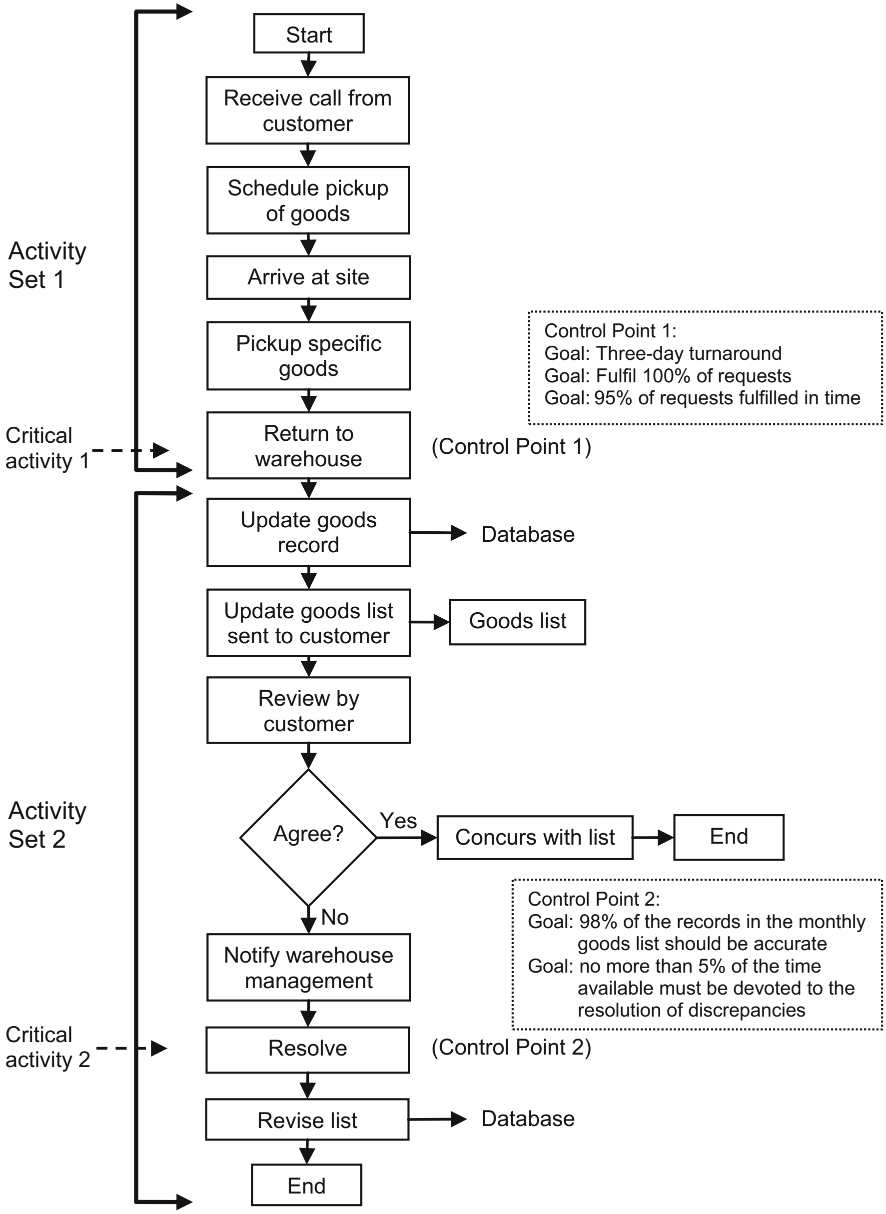

Fig. 5.20 Flow chart for the process of "goods pickup and storage", supplied with performance goals related to critical activities (U.S. Department of Energy – PBM SIG 2012). With permission

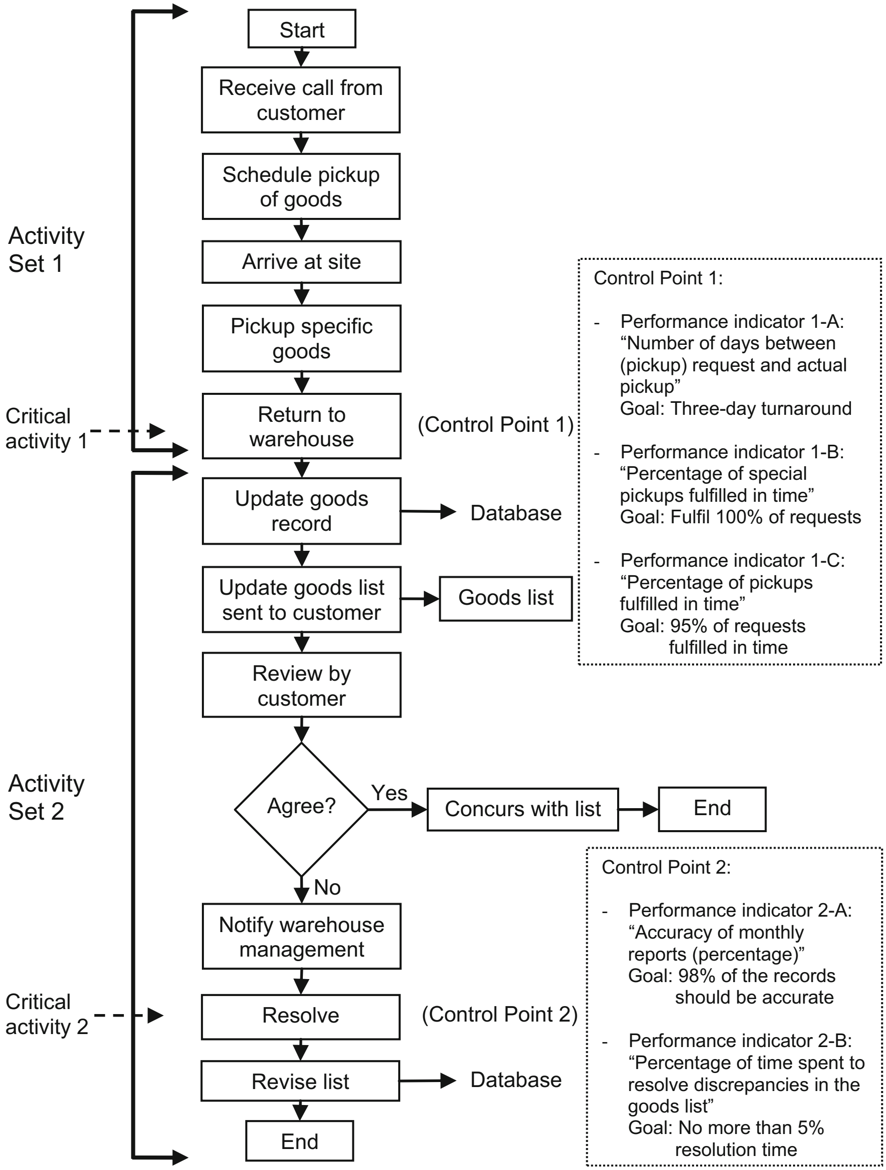

Fig. 5.21 Flow chart for the process of "goods pickup and storage", supplied with performance indicators selected (U.S. Department of Energy – PBM SIG 2012). With permission

$$\frac{\text{number of on time special pickups}}{\text{total number of special pickups}} \cdot 100 \tag{5.7}$$

- Collected data: "total number of special pickups that are scheduled in 1 week", distinguishing between those fulfilled in time and those not.
- Means of collection: goods transfer form.
- Frequency: weekly.

• *Performance indicator 1-C*: "Percentage of pickups fulfilled in time" (considering the totality of pickups):

$$\frac{\text{number of on time pickups}}{\text{total number of pickups}} \cdot 100 \tag{5.8}$$

- Collected data: total number of pickups that are scheduled in 1 week, distinguishing between those fulfilled in time and those delayed".
- Means of collection: goods transfer form.
- Frequency: weekly.

Considering the **critical activity 2** ("*Resolve discrepancies*"), we define the following indicators (see Fig. 5.21):

• *Performance indicator 2-A*: "Accuracy of monthly reports (percentage)":

$$\frac{\text{number of error free records}}{\text{total number of records}} \cdot 100 \tag{5.9}$$

- Collected data: "total number of records entries generated in each month", "number of errors detected" (used to calculate the "number of error-free records").
- Means of collection: goods list database, complaint logbook.
- Frequency: monthly.

• *Performance indicator 2-B*: "Percentage of time spent to resolve discrepancies in the goods list":

$$\frac{\text{total time spent for resolving discrepancies}}{\text{total time spent for managing goods list}} \cdot 100 \tag{5.10}$$

- Collected data: "time spent to resolve discrepancies in the goods list in each month"; "total time spent to manage the goods list in each month".

– Means of collection: estimate of the time spent to resolve discrepancies in the goods list (use of time cards).
– Frequency: monthly.

Step 5 *Identify responsible party(ies)*

The next step consists in identifying responsible parties for collecting data, analyzing/reporting actual performance, comparing it to goal/standard, and determining possible corrective actions.

In this specific case, two specialists are responsible for collecting, interpreting, and providing feedback on the data concerning goods. The warehouse supervisor is responsible for making decisions on possible improvement actions (see Fig. 5.22).

Step 6 *Collect data*

Data collection is much more than simply writing things down and then analyzing everything after a period of time. Even the best of measurement systems may fail because of poor data collection. Several preliminary analyses should be conducted in order to determine whether (1) the measurement system is working correctly, and (2) the frequency of data collection is appropriate.

In this specific case, two control points have been identified. The first one covers the activities 2, 4, and 5. The second one covers the activities 11 and 12 (Fig. 5.22).

For the first control point, the use of an existing goods transfer form is supposed to be the most efficient means for collecting the necessary information:

- for activity 2: the date in which the customer places the request and the scheduled date for pickup;
- for activities 4 and 5: the date in which the property is actually picked up and delivered to the warehouse.

Because of a variety in raw data, the data-collection approach at the second control point is somewhat more complex. The required information pieces are:

- for activity 11: a description of the problem and the date in which the division notified the warehouse (complaint logbook);
- for activity 12: a description of what is done to resolve the issue and the date action is taken (complaint logbook); the time spent by a property specialist in resolving the specific issue versus the total time spent on all work activities during the issue resolution period (time card records); the total number of reports distributed during the measurement interval (property reports).

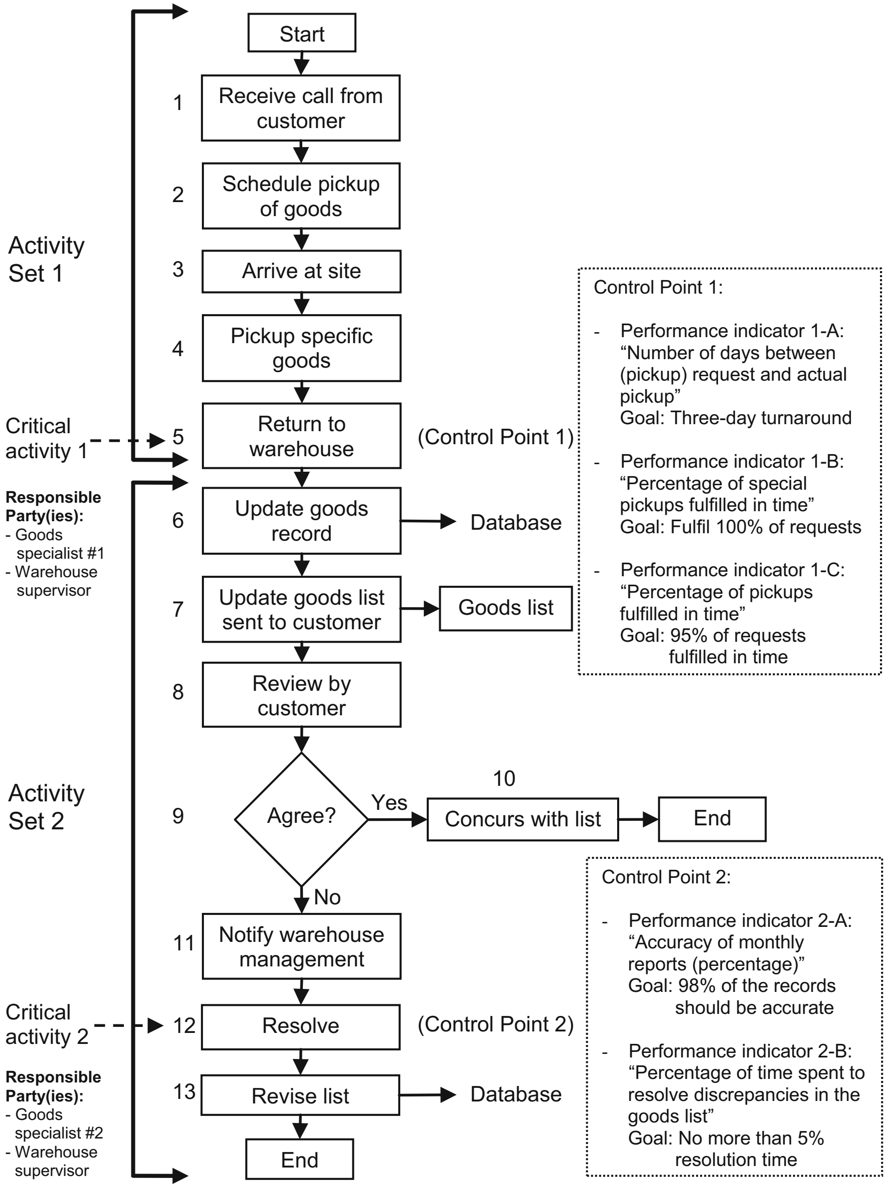

Fig. 5.22 Flow chart for the process of "pickup and storage of goods", supplied with the identification of the responsible parties for the two critical activities (U.S. Department of Energy – PBM SIG 2012). With permission

Step 7 *Analyze/report actual performance*

In this step, we will explore some of the classical tools to analyze and represent the results of performance indicators. A popular tool is represented by frequency charts depicting possible trends of indicators. Figures 5.23, 5.24, 5.25, 5.26, and

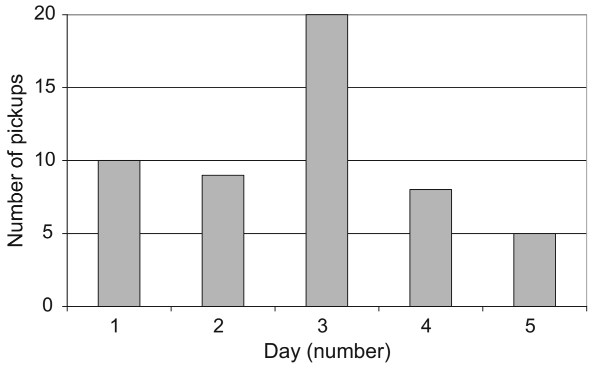

Fig. 5.23 Frequency chart related to indicator 1-A "Number of days between (pickup) request and actual pickup", with reference to the second week of a reference period

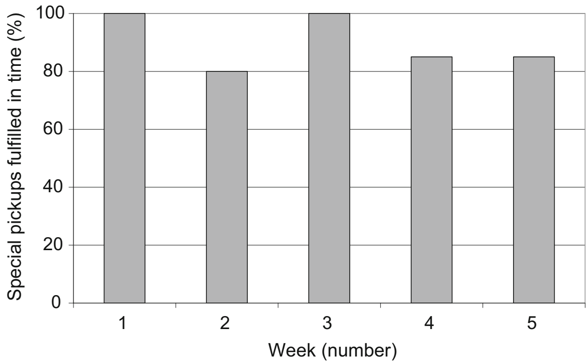

Fig. 5.24 Frequency chart related to indicator 1-B "Percentage of special pickups fulfilled in time", with reference to the first 5 weeks of a reference period

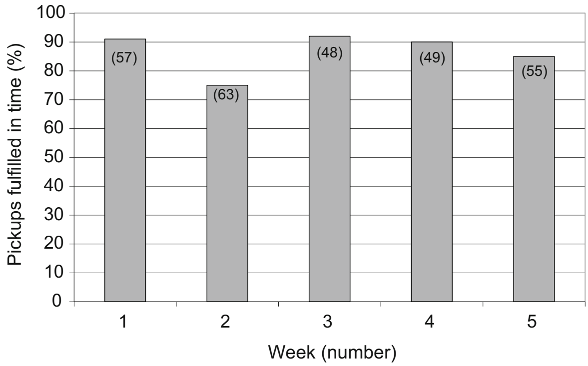

Fig. 5.25 Frequency chart related to indicator 1-C "Percentage of pickups fulfilled in time", with reference to the first 5 weeks of a reference period (total number of pickups in brackets)

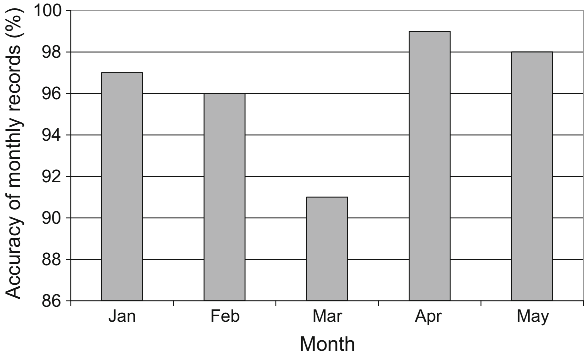

Fig. 5.26 Frequency chart related to indicator 2-A "Percentage accuracy of monthly reports", with reference to the first 5 months of a reference period

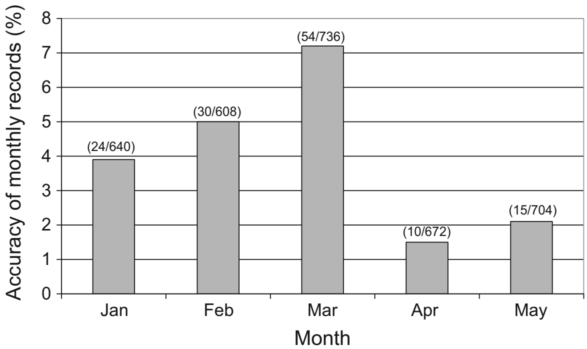

Fig. 5.27 Frequency chart related to indicator 2-B "Percentage of time spent to resolve discrepancies in the goods list", with reference to the first 5 months of a reference period (numerical values in brackets, corresponding to the "total time spent for resolving discrepancies" over the "total time spent for managing goods list")

5.27 show the frequency charts related to the five indicators defined at Step 4. Data refer to a 5-month monitoring period.

Step 8 *Compare actual performance with goals*

In this step, we compare actual performance to goals. Figures 5.23, 5.24, 5.25, 5.26, and 5.27 show that some objectives have not been met. For example, indicator 1-A ("number of days between pickup request and actual pickup") sistematically exceeds the target of 3 days. Consequently, we should determine whether gaps are significant and corrective actions are necessary.

Step 9 *Definition of corrective actions*

In this step, we need to take the necessary actions, in order to align the process with its objectives. Possible corrective actions are:

- Removing defects and their causes;
- Reducing the propensity of the process to generate defects;
- Enhancing the efficiency and effectiveness of the process.

For example, it can be noticed that the goal of 95% of (requested) pickups fulfilled in time is never achieved (see Fig. 5.25). As a consequence, it is necessary to find the causes of the problem and identify possible solutions.

5.6 Maintaining a Performance Measurement System

Performance indicators need to be constantly maintained and improved, in order to meet organizational targets. The basic elements of a performance measurement system are:

- strategic plan;
- key sub-processes;
- stakeholder expectations;
- organizational framework;
- regulations or standards;
- available technologies;
- employee involvement.

If some elements change, it is necessary to align the performance measurement system with them. For example, the presence of a new competitor or organitazional program may impact on stakeholder expectations; as a consequence, (part of) the performance measurement system may need to be changed.

5.7 (Mis)use of Indicators

Although indicators are instruments of undisputed utility, they can be misused, undermining the correct implementation of strategies (Winston 1993; Mintzeberg 1994; Perrin 1998). According to Perrin (1998), the most common types of misuse are:

- *Heterogeneous interpretations of the indicator meaning*

Indicators, independently on what they represent, are invariably used, recorded and interpreted in multiple ways. Thus, there can be a lack of comparability across different sites and staff, even when considering a simple indicator, such as the "number of clients served". For example, what is a "client" in a retail outlet? Is he/she anyone who phones or walks in the door to receive some information about products? Or is he/she someone who has a regular relationship with the outlet? If the same individual goes to different sales points of the same chain of shops, is he/she counted as one or multiple clients?

In general, indicators should be interpreted correctly, unambiguously and homogeneously. Training certainly contributes to obtaining homogeneous interpretations and encouraging commitment and sense of responsibility of workers; therefore it is particularly important for organizations with a high staff turnover.

Other possible misuses arise when workers/employees feel that their success may depend upon "making the numbers look good"; Perrin (1998) mentions an interesting example. Canada has employment equity legislation requiring federally regulated industries to develop and implement plans for equity in the employment of disadvantaged groups, including women, visible minorities, aboriginal people, and people with disabilities. One bank showed strong improvement in its record of employment of people with disabilities—until a legal centre representing the rights of people with disabilities grew suspicious and, upon investigation, discovered that the bank selectively changed its definition of disability to include a broader range of people, such as those requiring eyeglasses, as disabled, thus increasing their members.

- *Goal displacement*

When indicators become the objective, they result in "goal displacement", which leads to emphasis on the wrong activities, thus encourages means of "making the numbers" without improving actual outcomes. As a result, they frequently distort the direction of programs, diverting attention away from, rather than towards, what the program should be doing (cf. concept of *counter-productivity* in Sect. 4.6.1).

- *Use of meaningless and irrelevant indicators*

Not so unfrequently, indicators do not reflect reality. Perrin (1998) describes an example about some court house clerks who pointed out the difficulty in collecting the information required for their reports. When Perrin asked how they compile and submit their weekly statistics, he was told that: "We put down something that sounds reasonable".

Indicators can be irrelevant, even if they are accurate. The essence of a performance measurement system is to reduce a complex program to a small number of indicators (see Sect. 5.4). Indicators ignore the inherent complexity of social phenomena, which involve many interacting factors that cannot meaningfully be reduced to one or a limited number of quantitative indicators. In other terms, the problem is that of representing all the important dimensions of a process (cf. property of *exhaustiveness* in Sect. 4.6.2). There is an inverse relationship between the importance of an indicator and the ability to quantify it. As Patton (1997) indicated: "*To require goals to be clear, specific and measurable is to require programs to attempt only those things that social scientists know how to measure*". Many activities in the public policy realm are complex and *intangible* and cannot be reduced to numbers (Franceschini 2001). As Mintzberg (1996) stated: "*Assessment of many of the most common activities in government (or complex processes) requires soft judgement—something that hard measurement cannot provide.* [. . .] *Measurement often misses the point, sometimes causing awful distortions*". The above discussion emphasizes that what is measured often has little to do with what is really relevant.

- *Cost savings vs cost shifting*

Indicators typically look at individual processes, ignoring their inter-relationships. Consequently, "outcomes" may represent cost shifting rather than true cost savings, ignoring or transferring needs and clients elsewhere rather than actually addressing them. For the purpose of example, the number of "drop-outs" is a rather common indicator of the success of an University. A low number of drop-outs denotes the efficiency of University courses. Few systematic attempts, however, are made to discover why students leave, for how long and where they go. In some cases, students are merely transferred from a University course to one other; thus these students have not necessarily dropped out. In other words, the number of drop-outs is not necessarily an indicator of lack of effectiveness.

Indicators are invariably short-term in nature. But short-term benefits and outcomes may result in future requirements and increased costs over the longer term, thus "shifting" costs into the future.

- *Misuse of derived indicators*

Derived indicators may obscure subgroup differences. Consequently, the "same" outcome may reflect different forms of program effect. For example, the U.S. Census Bureau for 1996 showed an inflation-adjusted increase in income of 1.2% from the previous year. It indicates an improvement in the overall income of the people. However, a very different picture emerges if one examines the income of subgroups, revealing that for the wealthiest 20% the income had a 2.2% increase, while for the poorest 20% it had a 1.8% decrease. In other words, the "inequality gap" increased by a further 4%, in a year in which the economy was booming and unemployment falling. This finding would be totally missed by a derived indicator based on aggregate-family income.

• *Limitations of objective-based approaches to evaluation*

Three typical limitations of the use of objectives for evaluation purposes are:

– It is often difficult to combine the ambitiousness of objectives with their feasibility. Programs with ambitious objectives may be unfairly penalized, while mediocre programs are more likely to achieve their objectives;
– Objective-based approaches do not often take into account unintended or unanticipated consequences, which may be positive and negative;
– Objectives and relevant indicators are often fixed, forgetting that environmental conditions, needs and program activities are constantly changing. A responsive program should be able to adapt its objectives, trying to understand whether objectives are still desirable, or need to be adjusted. In fact, objectives and indicators may easily become out of date.

• *Useless for decision making and resource allocation*

A presumed role of performance indicators is to provide informed decision making and budgeting. But performance indicators are intrinsically useless for this purpose. As Newcomer (1997) states: "*Performance indicators typically tell what is occurring with respect to program outputs but they do not address how and why.*". For example, a program may fail to meet its performance targets because the program theory is wrong. But it also may fail to do so for a variety of other reasons, such as: inappropriate targets or indicators that do not identify important program outcomes; faulty management or implementation, under/over-funding, faulty statistics, etc. The use of indicators may sometimes incorrectly assume causality, i.e., inferring that the identified outcomes are a direct result of program activities. As evaluators know, causality can only be assessed through appropriate evaluations that aim at understanding the "whys", i.e., the mechanisms through which outcomes are achieved.

Thus, indicators provide no direct implications for action, unless other means are used to explore their (potential) impact on results. In general, making strategic decisions about future programs based upon indicators only can be dangerous and can lead to inappropriate action.

• *Less focus on outcome*

Perrin (1998) thinks that the sad, supreme irony is that performance measurement systems typically lead to less—rather than more—focus on outcome, innovation and improvement. A narrow focus on indicators is inconsistent with a focus on change and improvement that requires constant questioning about what else can be done or done better. The indicator misuse may lead to impaired performance, exaggerated emphasis on justifying and defending what was done, and reluctance to admit that improvement is needed.

Despite the problems and limitations identified above, performance indicators are indispensable for many activities like process evaluation, resources allocation, and comparison of complex systems. However, they should always be analysed, selected and used carefully.

5.8 Indicators as Conceptual Technologies

So far, we have considered the important role of indicators to represent a process. This section considers the role of indicators from another perspective: as anticipated in Sect. 1.1, indicators can influence organizations. In the following discussion, this concept is investigated in more depth.

Indicators can be considered as *conceptual technologies*, able to intangibly influence organizations (Barnetson and Cutright 2000). While the term "technology" typically refers to the technical procedures and instruments to generate products or services, the adjective "conceptual" refers to the notion of intangibility. Indicators can shape *what* issues we think about and *how* we think about those issues. This idea is derived from Polster and Newson's (1998) study concerning the external evaluation of the academic work. For example, a performance indicator that measures the employment rate of graduates indicates to institutions that this outcome is of importance to the agency that mandated its introduction; the act of measurement makes institutional performance on this indicator public.

In this way, the use of performance indicators shift the power of setting priorities and goals to those who control indicators themselves. In addition, indicators often represent a legitimisation of a specific policy. The use of indicators therefore affects the evolution of policies, since those who create and control indicators will have the power to determine what is really relevant. In addition, the use of indicators facilitates the use of financial rewards and punishments in order to manipulate institutional behaviour.

To examine the overall impact of a system of indicators, Barnetson and Cutright (2000) suggested the conceptual model in Table 5.25. This analysis of the impact of indicators is based on six dimensions: Value, Definition, Goal, Causality, Comparability and Normalcy.

By understanding the assumptions behind an indicator, it is possible to understand the relevant policy. Table 5.26 shows a set of questions to understand the potential impact of a *single* indicator. Table 5.27 shows a similar set of questions to understand the potential impact of a *set* of indicators.

It is generally accepted that performance indicators make knowledge transparent and quantitative (Porter 1995). According to Porter, quantification constrains the ability of others to exercise judgment when they use the information thereby subordinating personal bias to public standards. Such *mechanical objectivity* (i.e., following a set of rules to eliminate bias) is similar to the political and moral use of objectivity to mean impartiality and fairness. This differs from *absolute objectivity* (i.e., knowing objects as they really are) and *disciplinary objectivity* (i.e., reaching consensus with one's peers about the nature of objects).

Table 5.25 Six dimensions to evaluate the impact of indicators (Barnetson and Cutright 2000). With permission

Impact dimension	Description
Value	The act of measuring identifies the activities of value. In other terms, the inclusion or exclusion of indicators determines what is considered important or unimportant.
Definition	Performance indicators make it possible to analyse and interpret (part of) reality (e.g., accessibility, affordability, or quality of a service) by representing it.
Goal	Performance indicators include a reference point to evaluate the level of performance. This reference is also used for setting performance goals.
Causality	Performance indicators make it possible to assign responsibilities to activities or outcomes, according to a sort of relationship of causality.
Comparability	The use of "standard" indicators assumes that organizations are comparable. This may push organizations to modify their behaviour, in order to increase their performance.
Normalcy	Performance indicators determine expected results. This may push organizations to modify their behaviour, in order to increase their performance.

Table 5.26 Questions to understand the potential impact of a single indicator (Barnetson and Cutright 2000). With permission

Impact dimension	Questions
Value	What does this indicator (consider) as important to users?
Definition	How does this indicator turn a representation target into a measurable quantity?
Goal	Which results does the indicator encourage?
Causality	Who is responsible for the results represented through indicators? Which assumption of causality underlies this assignment of responsibility? For example, the fact that universities are responsible for the satisfaction of graduates entails that they can control/influence it somehow.
Comparability	In which way can organizations be compared? For example, measuring the generation of external revenue of colleges, universities and technical institutes implies that these institutions are somehow comparable.
Normalcy	Which assumptions does this indicator make about "normal" behaviours or outcomes? For example, measuring graduates' employment rate at a fixed point (after graduation) entails that graduates of all disciplines have similar career trajectories.

Indicators are therefore supposed to increase objectivity, through a set of shared rules. The application of indicators should increase impartiality of a decisional system, basing decisions upon facts rather than opinions. In general, organizational effectiveness tends to increase while increasing objectivity and linking resources to outcomes (Power 1996). This suggests that indicators are not mere technical means for evaluating performance but they are also policy tools (Barnetson and Cutright 2000). As policy tools, they may significantly reduce organizational autonomy (i.e., the freedom to make decisions).

Table 5.27 Questions to understand the potential impact of a set of indicators (Barnetson and Cutright 2000). With permission

Impact dimension	Questions
Value	Do indicators consider what is really important to users?
Definition	Do indicators operationalize representation targets, according to the major dimensions?
Goal	Are there trends in the goals assigned by this set of indicators? For example, do the indicators consistently reward organizations that reduce government costs by increasing efficiency?
Causality	Are responsibilities reasonably distributed among users or groups? Is there any causality trend behind this assignment of responsibility?
Comparability	Are indicators suitable to compare different organizations? For example, a set of indicators may consider (or ignore) differences in terms of goals, missions and resources.
Normalcy	Which activities and/or outcomes does the system of indicators assume to be normal?

The above considerations contradict the idea that the use of indicators determines more responsibilities. On the contrary, it somehow lead to confusing the concept of *responsibility* with that of *regulation* (Kells 1992). Since *regulation* requires an external authority that examines the performance of an organization, it therefore erodes autonomy rather than promoting it.

To study the consequences of the introduction of a performance measurement system, we consider the example of the Academic funding system in Alberta (Canada), which is based upon nine performance indicators (Barnetson and Cutright 2000). Five indicators are used by all institutions (i.e., the *learning* component) while four indicators affect only research universities (i.e., the *research* component). The total score of each institution is used to allocate funding (AECD 1999).

The five indicators related to the learning component fall into three categories based upon the government's goals of increasing *responsiveness*, *accessibility* and *affordability* (AECD 1997, 1999). Institutional responsiveness to the needs of learners and to provincial, social, economic and cultural needs are assessed by examining the *employment rates* of graduates and their *satisfaction* with their educational experience. Institutional progress towards higher levels of accessibility (i.e., increasing the number of enrolled students) is indicated by examining changes in full-load equivalent (FLE) enrolment, based on a 3-year rolling average. The affordability of institutions (i.e., the ability of providing learning opportunities to the greatest number of Albertans at a reasonable cost) can be deduced from the indicators "administrative expenditures" and "enterprise revenue".

Figure 5.28 shows the definition of indicators, while Table 5.28 shows their possible impact (only for *learning* component).

"Employment rate": percentage of graduate-survey respondents employed within a specified period (e.g., six months) after program completion

Points	0	15	20	25	30
Benchmarks	60%	70%	80%	90%	

"Graduate satisfaction with overall quality": percentage of graduate-survey respondents that are satisfied with the educational quality

Points	0	15	20	25	30
Benchmarks	70%	80%	90%	95%	

"Credit FLE enrolment": percentage variation in FLE credits from a period to the next one

Points		0	20	25	30
Benchmarks	Urban	-2%	0%	+4%	
	Rural	-5%	0%	+4%	

"Administrative expenditures": percentage with respect to total expenditures

Points		0	3	4	5
Benchmarks	>3500 students	11%	7%	5%	
	<3500 students	12%	8%	6%	

"Enterprise revenue": percentage with respect to government revenues

Points		1	3	4	5
Benchmarks	Urban	20%	35%	50%	
	Rural	10%	25%	40%	

Fig. 5.28 Indicators selected for evaluating the Academic funding system in Alberta (Canada). Indicators concern the learning component only (AECD 1997). With permission

As examplified, evaluating the impact of indicators is complex. The scheme of Barnetson and Cutright (2000) is an interesting contribute to this purpose. While physical systems are regulated by natural laws, independently from the way they are modelled, organizations are influenced by the way they are analysed and modelled (Hauser and Katz 1998).

Table 5.28 Analysis of the impact of the indicators for evaluating the Academic funding system in Alberta (Canada) (Barnetson and Cutright 2000). With permission

Indicator	Impact dimension	Specific description
Employment rate	Value	High levels of graduate employment are desirable.
	Definition	Responsiveness entails a good match between educational programs and real market needs.
	Goal	Institutions should increase the employment rate of graduates.
	Causality	Institutions can (1) control program offers and (2) match them with market needs.
	Comparability	Institutions are all able to generate some desirable outcomes for market.
	Normalcy	Graduates from different institutions have comparable career trajectories.
Graduate satisfaction	Value	High levels of graduate satisfaction are desirable.
	Definition	Responsiveness entails that educational programs satisfy graduates.
	Goal	Institutions should increase the satisfaction rate of graduates.
	Causality	Institutions can control the factors that contribute to the satisfaction of graduates.
	Comparability	Institutions are equally capable of satisfying their users.
	Normalcy	Graduates of different institutions have compatible program expectations.
Credit FLE enrolment	Value	Enrolment growth is desirable.
	Definition	Accessibility depends on the number of places available.
	Goal	Institutions should promote enrolment growth.
	Causality	Institutions can influence (1) the demand for places and (2) the availability of these places.
	Comparability	Institutions are equally able to increase enrolment.
	Normalcy	Economies of scale are equal between institutions.
Administrative expenditures	Value	Low levels of administrative expenditures are desirable.
	Definition	Affordability means minimizing administrative expenditures.
	Goal	Institutions should decrease administrative expenditures.
	Causality	Institutions can control the factors that contribute to administrative expenditures.
	Comparability	Institutions face similar economies (and diseconomies) of scale.
	Normalcy	Reduction of administrative expenditures per student, due to increased enrolment.
Enterprise revenue	Value	High levels of non-government/non-tuition revenue are desirable.
	Definition	Affordability entails maximizing external revenues.
	Goal	Institutions should increase generation of external revenue.

(continued)

Table 5.28 (continued)

Indicator	Impact dimension	Specific description
	Causality	Institutions can generate external revenue.
	Comparability	Institutions have similar abilities to generate external revenue.
	Normalcy	Raising revenue is compatible with the teaching function of institutions.

References

AECD. (1997). *Rewarding progress towards goals: Highlights of Alberta's performance envelope funding*. Edmonton: Advanced Education and Career Development.

AECD. (1999). *1998/99 Annual Report*. Edmonton: Advanced Education and Career Development.

Akao, V. (1988). *Quality function deployment*. Cambridge, MA: Productivity Press.

APQI (Associazione Premio Qualità Italia). (2018). *Migliorare le organizzazioni pubbliche attraverso l'autovalutazione*. Retrieved September 2018, from http://www.apqi.it/media/1128/caf_2013_italiano.pdf.

Atkinson, A. (1997). Linking performance measurement to strategy. *Journal of Strategic Performance Measurement, 1*(4), 5.

Auditor General of Canada and the Commissioner of the Environment and Sustainable Development. (2017). *Developing performance measures for sustainable development strategies*. Retrieved September 2018, from http://www.oag-bvg.gc.ca/internet/English/meth_gde_e_23715.html.

Barnetson, B., & Cutright, M. (2000). Performance indicators as conceptual technologies. *Higher Education, 40*, 277–292.

Bourne, M. C., & Bourne, P. (2011). *Handbook of corporate performance management*. Chichester: Wiley.

Brown, M. G. (1996). *Keeping score: Using the right metrics to drive world-class performance*. New York: Quality Resources, CRC Press.

DISPEA. (2004–2005). *Dipartimento di Sistemi di produzione ed economia dell'Azienda, Project: "Il nuovo Call Center del Comune di Torino: monitoraggio della qualità della comunicazione offerta ai cittadini"* (in alliance with AICQ and Comune di Torino).

DISPEA. (2005). *Dipartimento di Sistemi di produzione ed economia dell'Azienda, Project: "Valutazione della qualità nei servizi SO (Service Operations)"* (in alliance with RAI).

Eccles, R. (1991). The performance measurement manifesto. *Harvard Business Review, 69*(1), 131–137.

EFQM (European Foundation for Quality Management). (2013). Retrieved September 2018, from www.efqm.org.

Flapper, S. D. P., Fortuin, L., & Stoop, P. P. M. (1996). Toward consistent performance measurement systems. *International Journal of Operations and Production Management, 16*(7), 27–37.

Franceschini, F. (2001). *Dai Prodotti ai Servizi. Le nuove frontiere per la misura della qualità*. Torino: UTET Libreria.

Franceschini, F. (2002). *Advanced quality function deployment*. Boca Raton, FL: St. Lucie Press/CRC Press LLC.

Franceschini, F., Galetto, M., & Maisano, D. (2005). A short survey on air quality indicators: Properties, use, and (mis)use. *Management of Environmental Quality: An International Journal, 16*(5), 490–504.

Franceschini, F., Galetto, M., & Maisano, D. (2007). *Management by measurement: Designing key indicators and performance measurement systems*. Berlin: Springer Science & Business Media.

Franceschini, F., & Mastrogiacomo, L. (2018). Service gap deployment: A framework to link quality gaps to service activities. *International Journal of Services and Operations Management, 29*(1), 85–99.

Franceschini, F., & Romano, D. (1999). Control chart for linguistic variables: A method based on the use of linguistic quantifiers. *International Journal of Production Research, 37*(16), 3791–3801.

Hauser, J., & Katz, G. (1998). Metrics: You are what you measure! *European Management Journal, 16*(5), 517–528.

Houghton College. (2004). *Dashboard indicators: The top mission-critical indicators for Houghton College*. Retrieved September 2018, from http://campus.houghton.edu/.

Juran, J. M. (2016). *Quality control handbook* (7th ed.). New York: McGraw-Hill.

Kaplan, R. S., & Norton, D. (1992). The balanced scorecard-measures that drive performance. *Harvard Business Review, January–February*, 71–79.

Kaplan, R. S., & Norton, D. P. (2008). *The execution premium. Linking strategy to operations for competitive advantage*. Boston: Harvard Business School.

Kells, H. (1992). *Self-regulation in higher education: A multinational perspective on collaborative systems of quality assurance and control*. London: Jessica-Kingsley.

Lohman, C., Fortuin, L., & Wouters, M. (2004). Designing a performance measurement system: A case study. *European Journal of Operational Research, 156*, 267–286.

Mintzberg, H. (1994). *The rise and fall of strategic planning*. New York: The Free Press.

Mintzberg, H. (1996). Managing government, governing management. *Harvard Business Review, May–June*, 75–83.

Neely, A. (Ed.). (2002). *Business performance measurement: Theory and practice*. Cambridge: Cambridge University Press.

Neely, A., Richards, H., Mills, J., Platts, K., & Bourneal, M. (1997). Designing performance measures: A structured approach. *International Journal of Operations and Production Management, 17*(11), 1131–1152.

Nemhauser, G. L., & Wolsey, L. A. (1988). *Integer and combinatorial optimization*. New York: Wiley.

Newcomer, K. (Ed.). (1997). *Using performance measurement to improve public and nonprofit programs*, *New directions for evaluation* (Vol. 75). San Francisco: Jossey-Bass.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *The Journal of Marketing, 49*, 41–50.

Parker, R. G., & Rardin, R. L. (1988). *Discrete optimization*. San Diego: Academic Press.

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text*. Thousand Oaks, CA: Sage.

Perrin, B. (1998). Effective use and misuse of performance measurement. *American Journal of Evaluation, 19*(3), 367–379.

Polster, C., & Newson, J. (1998). Don't count your blessing: The social accomplishments of performance indicators'. In J. Currie & J. Newson (Eds.), *Universities and globalization: Critical perspective* (pp. 173–182). Thousand Oaks, CA: Sage.

Porter, T. M. (1995). *Trust in numbers: The pursuit of objectivity in sciences and public life*. Princeton, NJ: Princeton University Press.

Power, M. (1996). Making things auditable. *Accounting, Organizations and Society, 21*(2-3), 289–315.

Schmenner, R. W., & Vollmann, T. E. (1994). Performance measures: Gaps, false alarms and the usual suspects. *International Journal of Operations and Production Management, 14*(12), 58–69.

Thomson, J., & Varley, S. (1997). Developing a balanced scorecard at AT&T. *Journal of Strategic Performance Measurement, 1*(4), 14.

Thorburn, W. M. (1918). The myth of Occam's razor. *Mind, 27*, 345–353.

U.S. Department of Energy. (1996a). *Guidelines for performance measurement* (DOE G 120.1-5). Retrieved September 2018, from http://www.orau.gov.

U.S. Department of Energy. (1996b). *Critical few performance measures-definitions and methodology*. Office of Environmental Management (DOE/EM).

U.S. Department of Energy – PBM SIG (Performance-Based Management Special Interest Group). (2012). *Performance-based management handbook of techniques and tools–How to measure performance (revised)*. Retrieved September 2018, from http://www.orau.gov/pbm.

U.S. Department of Energy/Nevada Operations – DOE/NV. (1994, June). *Performance measurement process guidance document*. DOE/NV Family Quality Forum.

U.S. Department of the Treasury. (1994). *Criteria for developing performance measurement systems in the public sector*.

U.S. General Accounting Office. (2012). *Performance measurement and evaluation: Definitions and relationships (revised)*. Retrieved September 2018, from http://www.gao.gov/special.pubs/gg98026.pdf.

University of California. (2017). *SMART goals: A how to guide*. Retrieved September 2018, from https://www.ucop.edu.

Winston, J. (1993). Performance indicators: Do they perform? *Evaluation News and Comment, 2* (3), 22–29.

Index

F. Franceschini et al., *Designing Performance Measurement Systems*, Management for Professionals, https://doi.org/10.1007/978-3-030-01192-5