

The Complexity of Social Phenomena and the Construction of Indicators



Leonardo Salvatore Alaimo

1 Introduction

For many years, indicators have been considered a niche topic in the literature. In recent decades, this issue has become central to the scientific debate and has been discussed in any conference or workshop on the measurement and analysis of socioeconomic phenomena. Indicators are not a specific and exclusive topic of the natural or social sciences, but are used and constructed everywhere, and their functions in contemporary societies are widespread (Maggino et al., 2021).

To fully understand the importance of the concept of indicators in social sciences, their connection to the concepts of complex systems and measurement must be analysed and understood. Humanity has always had the need to know and understand reality and the phenomena defining it to achieve goals and satisfy needs and aspirations. Therefore, the need to generate knowledge is a defining feature in our lives. Consequently, the relationship between people and knowledge has always been a crucial topic in the reflection of scholars in every scientific discipline. Knowing reality refers to measuring reality. Measuring reality involves addressing complex systems and phenomena. Measuring complex phenomena involves dealing with indicators (Maggino & Alaimo, 2022). In the following pages, we try to describe these concepts.

L. S. Alaimo (✉)
Sapienza University of Rome, Rome, Italy
e-mail: leonardo.alaimo@uniroma1.it

2 Complexity and Complex Systems

2.1 Complexity: A Possible Definition

In recent decades, complexity has become a mainstream topic in different contexts and disciplines (e.g. physics, chemistry, biology, sociology, and psychology). The increasing attention to this concept coincided with the evolution of science, corresponding to the transition from *classical* to *modern* science (for details, see Alaimo, 2022). However, complexity in science has no precise meaning or unique definition (Érdi, 2008). As Morin (1985) states, the analysis of complexity cannot be addressed using a preliminary definition; there is no such thing as one complexity but different *complexities*. The influence of different disciplines on its conceptualisation has meant that this term has profoundly different meanings. Complexity does not belong to a particular theory or discipline, but rather to a *discourse about science* (Stengers, 1985). The term *complex* is often inappropriately used. We can understand its meaning by examining the differences from the concept of *complicated*, often used as a synonym, to refer to the difficulty in handling a situation or understanding a concept (Maggino & Alaimo, 2021). When dealing with particularly difficult situations or phenomena hard to explain, one tends to define them generically as ‘complex’ or ‘complicated’, giving these two concepts the same meaning. However, these two terms have profoundly different meanings, as reflected in their etymologies (De Toni & Comello, 2010; Letiche et al., 2012). ‘Complicated’ comes from the Latin *cum plicum*, in which the term *plicum* indicates the fold of a sheet. This term indicates something folded, which can be explained and understood by its unfolding. By contrast, ‘complex’ derives from the Latin *cum plexum*, where *plexum* means knot and weave. It refers to something woven, knotted, with interweaving, composed of many interconnected parts: ‘compound’ (Alaimo, 2021a, 2021b). Understanding a complicated phenomenon requires the adoption of an analytic approach; we must *unfold* the phenomenon in its creases and understand its basic components. Thus, understanding this phenomenon comes from understanding its components. It is always possible to achieve an understanding of a complex phenomenon, although this may seem difficult.

For instance, think of an embroidered tablecloth on a laid table and napkins that have the same embroidery, but it is not visible because they are folded. The embroidery on the latter will be immediately evident when we open them up by undoing the folds. The same thing happens when we try to solve a complicated problem: in order to understand it in its entirety (the embroidery hidden between the folds of the napkin), we have to identify its components (the folds of the napkin) and understand them (unfold them). (Maggino & Alaimo, 2022: 44)

Complex phenomena require a synthetic/systemic approach. We cannot understand the *plexum* simply by analysing its components; we must try to understand it as a whole.

Think of a nice jumper, with an intricate weave and many different colors. If we split up the jumper weave in its basic threads, we obtain a set of threads whose analysis does not help recreate the original system of the original jumper. In other words, if we consider the

Table 1 Main differences between complex and complicated

	Complex	Complicated
Etymology	<i>cum plexum</i> : something woven, knotted, with interweaving; composed of many interconnected parts; compound	<i>cum plicum</i> : indicates something folded, that can be explained and understood by unfolding its folds
Approach to knowledge	Synthetic/systemic approach: understand the phenomenon as a whole	Analytic approach: understand the individual components of the phenomenon

individual threads taken individually (adopting an analytic approach) we do not have a vision of the jumper, which comes from their interweaving. (Maggino & Alaimo, 2022: 44–45)

As Capra (1996) highlights, different approaches are necessary to understand complexity:

The properties of the parts can be understood only within the context of the larger whole. . . . Systems thinking is contextual, which is the opposite of analytical thinking. Analysis means taking something apart in order to understand it; systems thinking means putting it into the context of a larger whole. (Capra, 1996: 29–30)

The affirmation of the synthetic approach is one of the most important advances in twentieth-century science, closely linked to the awareness of understanding complexity using analysis:

Systems science shows that living systems cannot be understood by analysis. The properties of the parts are not intrinsic properties but can be understood only within the context of the larger whole. (Capra, 1996: 37)

A synthesis is not a reduction of reality but a stylisation highlighting the characteristics that arise from the interconnections among the elements defining a complex phenomenon. A complex phenomenon can sometimes be considered *difficult* because it cannot be explained. However, this difficulty does not depend on the complex nature of the phenomenon, but on the attempt to understand it using an analytical approach, merely breaking it down into its essential components rather than analysing it as a whole. We also need to clarify that a complex view of reality does not necessarily mean having a *complete* view of reality. The latter indicates that all components of a phenomenon are included with no missing data. However, having all the elements available and analysing them is not sufficient to understand a complex phenomenon. The latter can only be understood through the interconnections of the elements (Table 1).

2.2 *Complex Systems and Complex Adaptive Systems*

The word complex is often associated with *system*, a term used in common languages, and many scientific disciplines. Generally, a system can be defined as a set of elements that stand in interaction (Bertalanffy, 1968). More precisely, according

to Meadows (2009), it can be considered ‘an interconnected set of elements that is coherently organized in a way that achieves something’ (Meadows, 2009: 11). This definition highlights the main components of a system: *elements*, *interconnections*, and *functions*. A system is a collection of interconnected elements with a purpose. A system has its own behaviour, different from its parts, evolving over time according to changes that can concern the system and each of its essential components. Obviously, these changes could be shocking and unexpected. Most systems are able to withstand the impact of drastic changes thanks to one of their fundamental characteristics, *resilience*, that is the ‘system’s ability to survive and persist within a variable environment’ (Meadows, 2009: 76).

A system can be defined as an organic, global and organized entity, made up of many different parts, aimed at performing a certain function. If one removes a part of it, its nature and function are modified; the parts must have a specific architecture and their interaction makes the system behave differently from its parts. Systems evolve over time and most of them are resilient to change. (Alaimo, 2022: 21)

A complex system exhibits specific characteristics. It consists of a great variety of elements; this means that the elements are not only numerous, but also different from each other, making it difficult to understand. Moreover, elements are often other systems, which are in turn formed by systems, and so on. Complex systems are based on a *systemic hierarchy* that allows the control of elements, ensuring that they act in a coordinated and harmonious manner. They are ruled by what Haken (1983) defined as the *slaving principle*: the elements at a lower hierarchical level are slaves to the upper level and the overall system. In a complex system, the *interconnections* among elements are more important than the elements themselves. A high density and a variety of interconnections are typical. Complex systems consist of many different elements and relations, which can be analysed only in a synthetic way. In a complex system, elements and connections, besides being numerous, vary and differ. A particular category of complex systems is the so-called complex adaptive system (CAS), an open system consisting of various elements interacting with each other in a linear and non-linear way, which constitutes a unique and organic entity capable of evolving and adapting to the environment (Waldrop 1992). Holland (1992) underlined how all CASs share the same three characteristics: *evolution*, *aggregate behaviour*, and *anticipation*. They have the capacity to evolve and *learn*; they can adapt to the environment and change by processing information and building models capable of assessing whether adaptation is useful. Thus, they can survive.

As time goes on, the parts evolve in Darwinian fashion, attempting to improve the ability of their kind to survive in their interactions with the surrounding parts. This ability of the parts to adapt or learn is the pivotal characteristic of complex adaptive systems. (Holland, 1992: 19)

Complex adaptive systems present an aggregate behaviour that does not simply come from the behaviours of its elements, but *emerges* as a novelty from the interactions of the parts, as Morin (1977) affirms:

For the immune system, this aggregate behaviour is its ability to distinguish self from others. For an economy, it can range from the GNP to the overall network of supply and demand; for

ecology, it is usually taken to be the overall food web or the patterns of flow of energy and materials; for an embryo, it is the overall structure of the developing individual; for the brain, it is the overt behaviour it evokes and controls. (Holland, 1992: 19–20)

In addition to these two characteristics, there is a third that is difficult to understand: the typical ability of complex adaptive systems to anticipate changes. To adapt to changing circumstances, CASs develop rules that anticipate the consequences of certain responses. ‘At the simplest level, this is not much different from Pavlovian conditioning: “If the bell rings, then food will appear”’ (Holland, 1992: 20). Of course, the effects of such anticipation are complex, especially when a large number of elements are conditioned in different ways. Moreover, anticipation can cause large changes in aggregate behaviour, even when they do not come true.

‘The anticipation of an oil shortage, even if it never comes to pass, can cause a sharp rise in oil prices, and a sharp increase in attempts to find alternative energy sources’ (Holland, 1992: 20). Socioeconomic phenomena are CASs, consisting of a network of elements that interact with each other and with the environment. They are multidimensional and evolve by modifying their dimensions and the links between them. Therefore, knowledge of these phenomena must consider their complex nature. For this reason, measurements in the social sciences have typical characteristics that differ from those in the natural sciences. This requires the definition of systems of indicators capable of capturing the different aspects of the phenomena analysed. As can be easily understood, these systems are dynamic because they must adapt to changes in the measured phenomena.

The emergence of the concept of complexity has introduced many important innovations in the relationship between human beings and knowledge. In particular, the need for a new way of looking at reality emerges: the importance of going beyond empirical evidence and trying to grasp at the same time the whole and the individual components that compose it.

3 Measurement in the Social Sciences

Scientific knowledge is the result of a dialogue between logic and evidence, that is, it is generated from the interaction of two levels of scientific analysis: the *theoretical–formal level*, in which theories and hypotheses are developed and abstract concepts with their mutual relations are specified; and the *empirical level*, in which hypotheses are verified through empirical data (Maggino, 2017). Knowledge develops from the interaction, necessary and unavoidable, between the theory and observations realised by measurement. An empirical observation becomes a datum when evaluated within a theoretical framework. Thus, different types of data can be generated from the same empirical observations based on different theoretical frameworks, which are systems for comparing observations with one or more models. The relationship between the model and the observed data is the product of the measurement (Alaimo, 2022). If empirical observations are consistent with the model, it can

be concluded that the latter provides a good description of reality. Different models can represent reality with different levels of accuracy. At the same time, they are *falsifiable*; it is not possible to prove their truth because there is always a context in which a specific model can be inconsistent.

3.1 Measurement: Definitions and Main Aspects

The concept of measurement has an ancient origin. We can find the first definition of measurement in Book V of Euclid's Elements: measuring an attribute of an Object A means taking a reference Object B (called the unit of measurement) and determining how many times B is contained in A. Generally, measurement can be defined as the evaluation of the extension of a property in relation to a certain standard, the unit of measurement (Michell, 1999). Some attributes, such as velocity, height, and length, present a specific internal structure, namely, a quantitative structure. Consequently, these attributes were defined as quantities. Specific instances of a quantity are called the magnitudes of that quantity (e.g. the height of a person is the magnitude of the quantity, height). The magnitudes of a quantity are measurable because, based on the quantitative structure, they stand in relations/ratios to one another that can be expressed as numbers. A measurement can be defined as 'any method by which a unique and reciprocal correspondence is established between all or some of the magnitudes of a kind and all or some of the numbers, integral, rational, or real' (Russell, 2009: 176). This statement is the basis of the so-called representational approach, according to which 'measurement is the numerical representation of facts regarding the entities measured. A highly appreciated definition and a starting point for the reflections of other scholars is that of Stevens: measurement is the assignment of numerals to objects or events according to rules' (Stevens, 1946: 677). Based on Stevens' statement, for instance, Blalock (1968) defines measurement as a general process by which numbers are assigned to objects so that it is also understood which types of mathematical operations can be legitimately used. According to these definitions, measurement is an activity that determines a shift from the plane of reality in which we observe phenomena to the plane of numbers in which we try to encode them. This activity is meaningful and necessary. The rules of Stevens' definition must ensure that the translation is as faithful as possible so that any mathematical operations performed on objects are legitimate, as specified by Blalock. To ensure their meaningfulness, measures must be based on uniform procedures to collect, score, and report numerical results. In other words, they must be standardised. This ensures that possible foreign components representing the error of observation are isolated or minimised. Two types of error can be distinguished. The random error refers to all those factors that interfere with the measurement of any phenomenon and are ineradicable in the process. This type of error influences the reliability, that is, the consistency of a measurement model in terms of the degree of accuracy and precision with which the instrument measures and the ability to produce consistent measurements. The lower the random error, the higher the level

of reliability. The effects of such an error are completely systematic, and as a result, an instrument affected by it may overestimate or underestimate the magnitude of an attribute measured in a certain object. The systematic error determines the level of validity of the process, that is, the ability of a measurement procedure to measure what is intended to measure. There can be two types of systematic errors: methodological errors, that is, the error of definition/detection of the attribute to be observed, and the specific errors introduced by the observer in the use of the observation procedure. The lower the systematic error, the higher is the validity. Random error causes one measurement to differ slightly from the other because it is linked to unpredictable changes that occur during the process. The systematic error always affects measurements by the same amount or proportion, assuming a measurement is taken in the same way each time; it is predictable. Random errors cannot be eliminated; however, most systematic errors can be reduced. To reduce errors, all measurements must rely on a set of assumptions of different types (Alaimo, 2022):

- *Theoretical assumptions* related to the meanings given to the phenomenon measured.
- *Procedural assumptions* related to the rules of correspondence used in assigning numbers to observations.
- *Statistical assumptions* related to the main characteristics of statistical methods can be used for the analysis.

Compliance with these assumptions makes standardised measures.

3.2 Measurement in the Social Sciences: Systems of Indicators and Their Construction

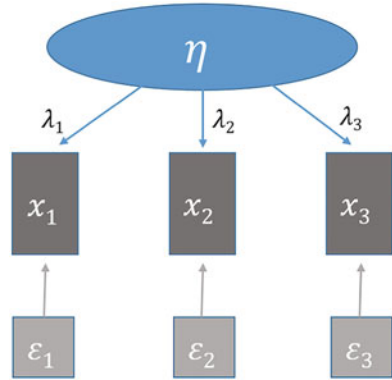
‘When social scientists use the term measurement it is in a much broader sense than the natural scientists do’ (Lazarsfeld, 1958: 100). With this statement from his well-known article “Evidence and Inference in Social Research” (1958), Lazarsfeld emphasises that in the social sciences, measurement has a typical character, which makes them not comparable to the natural sciences. The author made an essential contribution to the study and analysis of measurements in social sciences. He defined ‘operationalisation’ as the process through which theory and abstract concepts are translated into (measurable) variables. The variable is, therefore, the operationalised property of an object, since the concept to be operationalised must be applied to an object. ‘Between concept, property, and variable there is the same link that exists between the weight (concept), the weight of an object (property), and the weight of an object measured through the balance (variable)’ (Alaimo, 2022: 47–48).

Measurement in the social sciences is influenced by objects. Socioeconomic phenomena are complex adaptive systems, and, consequently, the approach to understanding them must take into account their nature. Measuring these phenomena means trying to understand their nature, understanding each of them as a whole. In

this field, dealing with measurements means dealing with systems of indicators. What is an indicator? This can be considered as the result of the translation of reality to the plane of numbers. The term is often used synonymously with an index, but its meanings are profoundly different. The meaning of the term index is *anything useful to indicate*, and it is used in statistics with multiple meanings. The indicator is what relates concepts to reality (Maggino, 2017: 92). Horn (1993) defined indicators as purposeful statistics. An index becomes an indicator only when its definition and measurement occur within the ambit of a conceptual model. Given the complex and multidimensional nature of socioeconomic phenomena, their analysis involves the identification of different basic indicators connected in a system. Each indicator constitutes what is currently measured, with reference to a specific aspect or dimension of a phenomenon. A system of indicators is not a simple collection of measures, but a complex system. Indicators within a system are interconnected, and new properties typical of the system emerge from these interconnections. The development of systems of indicators must strictly follow a set of rules codified in a step-by-step process, the so-called hierarchical design (Maggino, 2017), which is a specification of Lazarsfeld's operationalisation. The starting question is, what is the phenomenon to be studied? Defining a phenomenon is not an easy task, based on a process of abstraction influenced by different factors, such as the sociocultural and spatial-temporal context in which the phenomenon is studied. Consequently, various definitions are possible and legitimate. Indeed, the definition of phenomena is subjective because it always depends on the researchers' point of view, on the *small windows* through which they observe reality and make hypotheses on it. Evidently, it is necessary to prevent this subjectivity from becoming arbitrary, involving no relationship with reality. The second step is the identification of latent variables, each of which is an aspect to be observed. These reflect the nature of the phenomenon, which is consistent with the conceptual model. Based on its level of complexity, a variable can be described by one or more factors. The different factors of each variable are referred to as dimensions. This concept is complex and theoretical. It is possible to handle profoundly different situations. The latent variable can assume only one underlying dimension. In other situations, we can deal with latent variables with two or more dimensions. Once the latent variables and their dimensionality are identified, the next phase consists of the selection of basic indicators. We can adopt a single indicator approach by measuring each latent variable using a single indicator. This approach could be weak because it is based on the assumption of direct correspondence between one latent variable and one indicator. Generally, the multi-indicator approach is preferable, in which, for each latent variable, several indicators are identified and selected. This approach increases measurement accuracy and precision, compensating for random errors.

The rigorous application of hierarchical design and adherence to its underlying assumptions enables the creation of a system of indicators suitable for measuring a particular phenomenon. One of the main assumptions concerns the specification of the model of measurement. The measurement model describes the relationship between a construct and its indicators. We can deal with two models: the *reflective* and the *formative* (Curtis and Jackson, 1962; Blalock, 1964; Diamantopoulos &

Fig. 1 Reflective measurement model: An example with three indicators and one latent variable



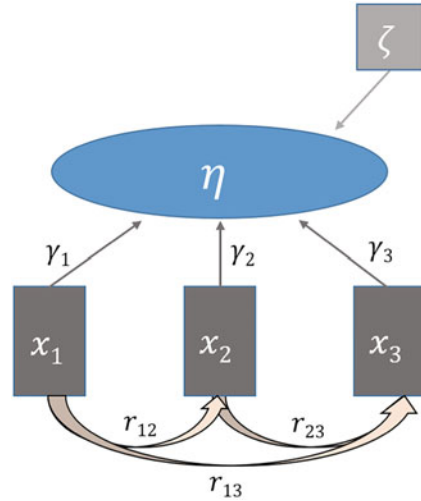
Siguaw, 2006; Diamantopoulos et al., 2008). In reflective measurement models, causality is from the construct to the measures, that is, measures are considered the effects of an underlying latent construct (Bollen & Lennox, 1991). The following equation explains this relationship:

$$x_i = \lambda_i \eta + \varepsilon_i$$

where x_i is the i -th indicator, η is the latent variable, λ_i is the coefficient capturing the effect of the latent variable on the i -th indicator, and ε_i is the measurement error for the i -th indicator. Figure 1 summarises the main components of the reflective model.

In this model, indicators reflect the latent variable and correspond to the linear functions of the underlying construct and measurement error. Each indicator has a specific error term, assumed to be mutually independent ($\text{cov}[\varepsilon_i, \varepsilon_j] = 0$ for $i \neq j$) and unrelated to the latent variable ($\text{cov}[\varepsilon_i, \eta] = 0 \forall i$). Thus, changes in the latent variable cause variations in all indicators simultaneously, and all indicators must be positively correlated. Internal consistency is fundamental: correlations between indicators are explained by the measurement model, and two uncorrelated indicators cannot measure the same construct (Bollen, 1984). This model is typical in psychometric research, such as in the measurement of attitudes. ‘Let’s suppose we want to measure the intelligence of a group of individuals using the results obtained by each of them in a series of tests. In this hypothesis, it is quite evident that the intelligence of each individual influences the result of the tests and not vice versa. As a consequence, we expect that the results of an individual to the different tests are quite the same and, from a statistical point of view, correlated with each other (because they are determined by the same latent variable). If a test gives a completely different result, it does not measure that specific construct’ (Alaimo, 2022: 55–56). Formative models typically measure socioeconomic phenomena in which indicators cause a latent variable (Curtis & Jackson, 1962; Land, 1970). ‘Let’s suppose we want to measure the gender inequality. We must start with its definition: we can say that it refers to systematic differences in the outcome of men and women on a variety of issues ranging from economic participation and opportunity, political

Fig. 2 Formative measurement model: An example with three indicators and one latent variable



empowerment, and educational attainment to health and well-being. In this case, by means of the definition, we already identify the components that form the concept and, consequently, the indicators to be selected. According to this definition, a measure of the gender inequality must take into account economic participation and opportunity, political empowerment, and educational attainment to health and well-being and use at least one indicator to measure each of them. If one of these dimensions is not taken into account, the concept of gender gap changes' (Alaimo, 2022: 58). Figure 2 shows the main components of the formative models.

The model is specified by the following equation:

$$\eta = \sum_{i=1}^n \gamma_i x_i + \zeta$$

where x_i is the i -th indicator, η is the latent variable, γ_i is the coefficient capturing the effect of the i -th indicator on the latent variable, and ζ is the measurement error that includes all remaining causes of the construct not represented in and not correlated to the indicators ($\text{cov}[x_i, \zeta] = 0$). Indicators do not present specific measurement error terms (Edwards and Bagozzi, 2000). According to this model, indicators are not replaceable; thus, changing an indicator will change the construct. Correlations among indicators are not explained by the measurement model, and internal consistency is of minimal importance; formative indicators might correlate positively or negatively, or lack any correlation (Bollen, 1984). There is a heated debate in the literature on the use of these two models. In particular, authoritative scholars have strongly criticised and opposed the use of formative measurement models (Howell et al., 2007; Wilcox et al., 2008; Edwards, 2011). Other scholars have strongly supported the effectiveness of formative models (Bollen, 2007; Diamantopoulos et al., 2008; Bollen & Diamantopoulos, 2017). The debate in the literature continues

to be animated, and it is not the aim of this paper to report this in detail. It is important to clarify that the choice of the measurement model does not depend directly on the researcher, but only on its appropriateness to the phenomenon that one intends to study. If the direction of the relationship is from the construct to the measures, we have a reflective model: by contrast, if the direction of the relationship is from the measures to the construct, we have a formative model (Coltman et al., 2008; Alaimo, 2022).

A system of indicators is a complex system, the analysis and understanding of which require approaches that allow more concise views. As Lazarsfeld (1958) states, the concept needs to be reconstituted, and all indicators within the system must be brought back to a synthesis. Synthesising data responds to a range of cognitive and practical needs, which is justified by the fact that knowledge of complex phenomena involves some form of *reductio ad unum* (Sacconaghi, 2017). From a methodological point of view, synthesis can concern different aspects of a multi-indicator system (Maggino, 2017):

- The synthesis of *statistical units* aims to aggregate the units in order to create macro-units for comparison, with reference to the indicators of interest. The statistical methods that allow for this are part of the *cluster analysis*. In this chapter, we will not dwell on these techniques, the literature of which is vast and deserves a separate discussion (for more information about cluster analysis, see Landau et al., 2011; Hennig et al., 2015; Maharaj et al., 2019).
- The synthesis of *statistical indicators* aims to aggregate the values referring to several indicators for each unit of observation, obtaining a synthetic measure. From a technical point of view, the statistical methods used in this case can belong to two different approaches: aggregative–compensative and nonaggregative.

Obviously, these two aspects are not mutually exclusive; however, it is often necessary to do both for a full understanding of reality (Alaimo, 2022). This chapter focuses on the synthesis of statistical indicators.

3.3 *Synthesis of Multi-indicators Systems*

As pointed out previously, the complex and multidimensional nature of socioeconomic phenomena requires the adoption of different measures to analyse and understand them. The measurement process in the social sciences is associated with the construction of systems of indicators, which makes it possible to measure phenomena that would not otherwise be measurable. Similar to the phenomena they must measure, these systems are also complex adaptive systems. The complex nature of such systems requires a synthetic approach to understand the phenomena as a whole. This implies the use of various basic indicators and criteria for summarising them. A basic indicator can be defined as an indirect measure of a phenomenon that cannot be directly measured. From this perspective, an indicator is not simply raw

statistical information, but represents a measure organically linked to a conceptual model aimed at describing different aspects of reality. It can be defined as a constructed variable related to a specific aspect or dimension of a complex phenomenon. Synthetic indicators are obtained by properly synthesising elementary indicators according to established criteria and rules. It is right to emphasise the adverb properly: in fact, if the construction of a synthetic index is not done according to specific steps and rules (i.e. properly), the resulting measure may inadequately represent reality and lead to misleading conclusions. Synthetic indicators have been widely used in the literature and various fields. The main purpose of their success is informative. It is easier for the public to understand a synthetic indicator (a single measure) than many elementary indicators.

Before analysing the main methods for synthesising multi-indicator systems in detail, it is necessary to formalise them mathematically. Generally, they consist of a set of measures (the basic indicators) at different measurement scale levels, observed on a set of statistical units. In its simplest form, a system of indicators is a matrix of data X typical of multivariate statistics:

$$X \equiv \{x_{ij} : i = 1, \dots, N; j = 1, \dots, M\} \equiv \begin{pmatrix} x_{11} & \cdots & x_{1M} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{NM} \end{pmatrix}$$

where the $i = 1, \dots, N$ rows represent the statistical units, the $j = 1, \dots, M$ columns represent the indicators, and the generic unit x_{ij} represents the determination of the j -th indicator in the i -th unit. We must clarify that in this study, we consider the simplest formalisation of the synthesis question, in which we do not deal with the temporal dimension. Indeed, in most cases, the multi-indicator systems are in the form of *three-way data time arrays* of type ‘same objects \times same variables \times times’, algebraically formalised as follows:

$$Y \equiv \{y_{ijt} : i = 1, \dots, N; j = 1, \dots, M; t = 1, \dots, T\}$$

where indices i, j , and t indicate the units, indicators, and times, respectively, and x_{ijt} is the value of the j -th indicator observed in the i -th unit at time t -th. These data structures are characterised by a greater complexity of information, consisting of the fact that multivariate data are observed at different times (D’Urso, 2000). In this chapter, we chose not to deal with the synthesis of three-way data time arrays, the complexity of which requires deeper knowledge of the subject (for an overview of the main synthetic methods for three-way data time arrays, see, e.g. Alaimo (2022)).

Given the bi-dimensional data matrix X , the goal of the synthesis is to obtain a vector $v \equiv \{v_i\}$ with N statistical units, in which the generic element v_i represents the synthetic value of the i -th unit with respect to all the J indicators:

$$X \equiv \begin{pmatrix} x_{11} & \cdots & x_{1M} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{NM} \end{pmatrix} \mathbb{Q} \mathcal{V} \equiv \{v_i\} \equiv \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix}$$

Focusing on how to obtain the synthesis of indicators from a technical point of view means focusing on the arrow \mathbb{Q} of the previous equation. In the literature, there are two different approaches to synthesis: aggregative-compensative, and non-aggregative. It should be clarified that one approach is not better than the other; each has pros and cons, and their use also (and especially) depends on the nature of the indicators. This is a crucial point. As clarified in the previous pages, indicators within a system can belong to different levels of the scale of measurement (Stevens, 1946). This is a relevant issue because the properties of the indicator determine the type of statistical tool that can be used to study it, and consequently, influence the choice of method of synthesis for a system of indicators. However, this issue is often underestimated. The aggregative-compensative approach is the dominant framework in the literature. As the name suggests, it consists of the mathematical combination (or aggregation) of a set of indicators by applying methodologies known as composite indicators (Saisana & Tarantola, 2002; Nardo et al., 2005; OECD, 2008). It is evident that the assumption underlying the construction of a composite is the possibility that the basic indicators are mathematically combinable and therefore cardinal. Despite such evidence, in the literature, several studies deal with nominal or ordinal indicators as if they were cardinal, using for their synthesis tools that are inappropriate to their level of scale (for instance, the arithmetic or geometric mean). Over the years, research has focused on identifying methods suitable for dealing with systems of indicators at different scaling levels. Thus, the so-called non-aggregative approach gradually became widespread: the synthetic indicator was obtained without any aggregation of the basic indicators. Different methodologies have been proposed within this approach, such as social choice theory (Sen, 1977; McLean, 1990, Arrow, 2012) or multi-criteria analysis (Nijkamp & van Delft, 1977; Macoun & Prabhu, 1999; Belton & Stewart, 2002; Ehrgott et al., 2005; Zopounidis & Pardalos, 2010). In particular, the partially ordered set (poset) theory (Neggers & Kim, 1998; Schroder, 2002) has become a reference, as evidenced by the large number of studies using this method for both ordinal (see, for instance, Fattore, 2016, Alaimo et al., 2022b, 2023, Fattore & Alaimo, 2023) and mixed (see Bruggemann & Patil, 2011; Kerber, 2017; Alaimo et al., 2021a, 2021b, 2022a) indicator systems. In the following pages, we focus on the aggregative-compensative approach and on systems in which all indicators are cardinal.

3.4 The Aggregative-Compensative Approach

As specified previously, the aggregative-compensative approach involves the aggregation using a mathematical function of the basic indicators. Therefore, a composite indicator is a measure based on sub-indicators that have no common meaningful unit

of measurement, and there is no shared method of weighting these sub-indicators. Synthesis is a measure not necessarily a number. This can be an image, as highlighted by the literature on the use of metaphoric images for the representation of phenomena (Tufté, 2001; Lima, 2013). Some authors (for instance, Diener & Suh, 1997) have criticised the choice of constructing a single composite index, pointing out that a more appropriate choice would be to use a dashboard. This is an open issue in the literature, and we can find arguments supporting composites or against them. A dashboard allows one to avoid an arbitrary choice of the functional form and weighting scheme and to observe a phenomenon from multiple points of view. However, this does not allow for a simple and direct understanding of the phenomenon under consideration. Constructing a composite is not an easy task and involves the implementation of different steps and a series of decisions and choices: the selection of basic indicators, whether and how to normalise them, and which aggregation procedure to choose. Although guided by knowledge of the phenomenon, most of these choices are subjective and, therefore, often considered non-scientific. This is one reason composite indicators have been considered a niche field in the literature for many years. Beyond these critics, composites are widely disseminated and used in the scientific literature and policymakers. We must clarify that there is no universal method for the construction of composites that must be guided by expert knowledge of the phenomenon.

The construction of a composite indicator is a step-by-step process (Nardo et al., 2005; OECD, 2008):

- Definition of the phenomenon
- Selection of the basic indicators
- Exploratory analysis of basic indicators
- Normalisation of individual indicators
- Aggregation of the normalised indicators
- Index validation

The steps are hierarchically ordered; therefore, the next step presupposes the previous step. The first two steps are theoretical, but they are not considered separate from the statistical-methodological ones (the other three).

In the previous pages, we discussed that measurement in the social sciences begins with the definition of the phenomenon. The concept must always be referred to and inserted within a theoretical framework that provides meaning. Particular attention should be given to the measurement model as we have seen in the previous pages. The choice of the measurement model depends on the appropriateness of the phenomenon to be measured and on the nature and direction of the relationships between constructs and measures (Alaimo, 2022). All socioeconomic phenomena require a formative measurement model. Therefore, in the following pages, we assumed that we deal with formative measurement model. The reflective measurement model is most widely used in the psychological and management sciences. The synthetic approaches and methods that allow us to deal with reflective models differ from those typical of the formative. One of the main methods in reflective models is undoubtedly factor analysis (Spearman, 1904; Thurstone, 1931; Cattell, 1978). It

must be clear what the composite wants to measure. If a phenomenon is poorly defined, it will certainly be poorly measured. However, the opposite was not true. If the phenomenon is well-defined and the matrix is composed of indicators of good quality, it is not necessarily the case that the composite index is valid (e.g. if the methodology used is not consistent with the indicators).

The selection of indicators is a delicate step that cannot be conducted independently of the others. The choice of basic indicators is based on a theoretical framework. Therefore, the approach used is based on a reasoned selection of the indicators included in the system. One question that must be addressed is, how many indicators should we consider? There are no unequivocal answers to this question. The general rule is that all dimensions of the phenomenon must be represented and measured using at least one indicator. Consequently, each latent variable can be defined and measured by using a single indicator. This single indicator approach is weak and assumes the existence of direct correspondence between one latent variable and one indicator. It is preferable to adopt a multi-indicator approach, that is, using several indicators for each dimension. This approach allows the overcoming (or at least reduction) of problems produced by the single indicator approach. In fact, using multiple indicators increases the measurement accuracy and precision, allowing one to compensate for random errors. Simultaneously, the risk is that the indicators are redundant. Redundancy can be defined as the excess of significant elements and information compared to what is strictly necessary for the correct understanding of a message. It is often intentional to increase the probability of complete reception of the message, even in the presence of noise or disturbances. The redundancy of indicators in a system can be useful in increasing the reliability of the measurement; the multi-indicator approach reduces the random error. However, we often encounter systems with too many indicators in which synthesis may not be significant or even possible. Therefore, it was necessary to reduce the number of indicators. There is not always a valid rule for this choice that should always be made with the theoretical framework and measurement model in mind. Dealing with a reflective measurement model, if it is necessary to eliminate indicators from the system, we can begin with those that are not correlated with the others because they do not measure the latent reflective variable considered. But even if we eliminated one indicator correlated with the others, we would have no change in the latent variable which 'causes' the indicators and remains unchanged. However, the formative models are different. The exclusion of an elementary indicator always affects the latent variable and, consequently, the composite indicator. This is because the indicators 'cause' the latent variable and remove (or add) one change (perhaps even slightly) the latent variable. Moreover, if we wanted or needed to eliminate an indicator, it would be more appropriate to act on indicators that are highly correlated with each other rather than to eliminate an indicator not correlated with the others and that, consequently, measures a different aspect of the phenomenon. In general, we need to choose a number of indicators that allow us to adequately represent the desired conceptual dimension, avoiding redundancy and ensuring the reduction of error by finding a compromise between possible redundancies caused by overlapping information and the risk of losing information (Salzman, 2003).

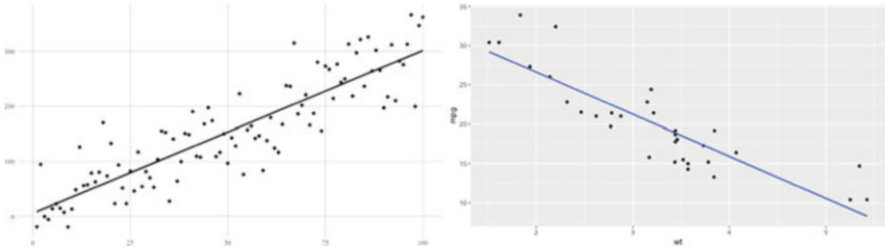


Fig. 3 Examples of linear correlation

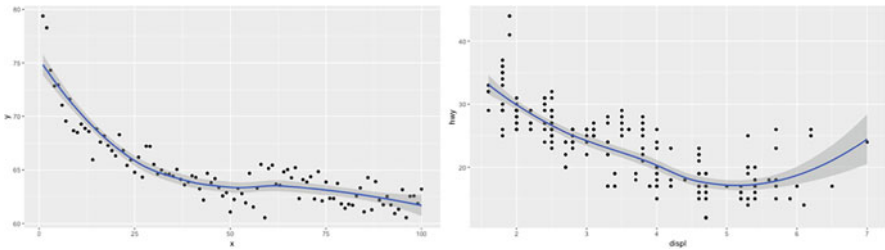


Fig. 4 Examples of non-linear correlation

The exploratory analysis of basic indicators is an important methodological step that aims to answer important questions. Is the latent structure of the synthetic index well defined? Are the chosen indicators sufficient to describe the phenomenon? It involves the application of multivariate statistical techniques to study the latent structure of data and analyse the relationships among the indicators within the system. The traditional approach involves the study of correlations between elementary indicators and principal component analysis (PCA). The term correlation in statistics indicates a reciprocal relationship between phenomena; in particular, it refers to the reciprocal relationship between two quantitative characteristics. Given two quantitative characters, X and Y , there is a positive correlation or concordance between them when they tend to increase or decrease together; in other words, when as one increases (or decreases), so does the other. There is a negative correlation or discordance when; as one variable increases, the other tends to decrease. Correlation is a symmetrical concept that does not refer to a cause-and-effect link but to the tendency of one variable to change in relation to another. When discussing the correlation, two aspects must be considered: the type of relationship between the two variables and the form of the relationship. The relationship can be linear if (in extreme simplicity) one graphically represents the double distribution through a scatter plot, the cloud of points approximates a straight line, as in the examples reported in Fig. 3.

There is a non-linear correlation if one by graphically represents the double distribution through a scatter plot, the cloud of points will have a non-linear (curvilinear) trend, as in the examples reported in Fig. 4.

Regarding the form of the relationship, we need to consider the direction, which can be positive (if as one variable increases, so does the other) or negative (if as one variable increases, the other decreases), and the magnitude, which refers to the strength of the relationship between the variables. Correlation coefficients are used to express the relationship between two variables in terms of both magnitude and direction. The correlation coefficient takes values within the range $[-1, 1]$:

$$-1 \leq \phi \leq +1$$

- The maximum value 1 in the case of perfect positive correlation
- The minimum value -1 in case of perfect negative correlation
- the value 0 in case of uncorrelation.

For exploratory analysis, the most commonly used coefficients for analysing the correlation between two variables X and Y are as follows:

1. Pearson correlation coefficient $r_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X\sigma_Y}$
 where $\text{cov}(X, Y)$ are the covariances, σ_X is the standard deviation of X , and σ_Y is the standard deviation of Y .
2. Spearman's rank correlation coefficient $\rho_{X,Y} = 1 - \frac{6 \sum_{i=1}^N d_i^2}{N(N^2 - 1)}$
 where $d_i = r(x_i) - r(y_i)$ is the difference between the two ranks of the i -th observation.
3. Kendall rank correlation coefficient $\tau = \frac{(c - d)}{(c + d)}$
 where c is the number of concordant pairs and d is the number of discordant pairs.

Although important, correlations are not decisive; in the context of constructing synthetic indicators, they can be considered as a guide. The first thing to consider is the measurement model, remembering that it depends not on an arbitrary choice of the researcher but on the definition of the phenomenon and the consequent nature of the latent variable. The importance of studying correlations is evident in the case of a reflexive measurement model. In fact, the indicators, in this case, are a 'reflex' of the latent variable. Thus, the correlation between the indicators is explained by the measurement model, and the two uncorrelated indicators cannot measure the same latent variable. Therefore, correlation analysis allows for the exclusion of indicators unrelated to the latent variable. In the case of formative models, the study of correlations is equally important. In this case, the internal consistency of the indicators is of minimal importance, and two unrelated indicators can be relevant to the same construct. Simultaneously, two highly correlated indicators are likely to measure the same aspect of the phenomenon (redundancy). PCA is a multivariate statistical technique used in the composite indicator field for various purposes:

- To identify the dimensionality of the phenomenon
- To define the weights
- As an aggregation method

This technique was first described by Karl Pearson (1901), and was later independently developed and named by Harold Hotelling (1933). Let us consider data matrix X with N statistical units and M cardinal indicators, as previously described. The aim of PCA is to take the M variables V_1, \dots, V_M and find linear combinations of these to produce principal components Z_1, \dots, Z_M that are uncorrelated:

$$Z_j = \sum_{i=1}^M a_{ij} V_i \quad j = 1, 2, \dots, M$$

The weights a_{ij} are chosen such that the principal components Z satisfy the following conditions:

- They are uncorrelated (orthogonal).
- The first principal component accounts for the maximum possible proportion of the variance of the set of original variables, the second principal component accounts for the maximum of the remaining variance, and so on, until the last component absorbs all the remaining variance that is not accounted for by the preceding components.

$$a_{1j} + a_{2j} + \dots + a_{Mj} = 1 \quad j = 1, 2, \dots, M$$

PCA just involves finding the eigenvalues λ_j of the covariance matrix C :

$$C = \begin{bmatrix} c_{11} & \cdots & c_{1M} \\ \vdots & \ddots & \vdots \\ c_{M1} & \cdots & c_{MM} \end{bmatrix}$$

where the diagonal element c_{ii} is the variance of V_i and c_{ij} is the covariance of variables V_i and V_j . The eigenvalues of matrix C are the variances in the principal components. There were M eigenvalues. Negative eigenvalues are not possible in a covariance matrix. An important property is that the sum of the variances of the principal components is equal to the sum of the variances of the original variables.

$$\lambda_1 + \lambda_2 + \dots + \lambda_M = c_{11} + c_{22} + \dots + c_{MM}$$

Before performing PCA, the original variables were commonly standardised to have zero means and unit variances to avoid one variable having an undue influence on the principal components. Thus, matrix C takes the form of a correlation matrix. In this case, the sum of the diagonal terms, and hence the sum of the eigenvalues, is equal to M , which is the number of variables. The correlation coefficients of the principal components Z with the variables V are defined *loadings*, r_{Z_j, V_i} (for a more

Table 2 Example: System of three cardinal indicators observed in four units

	V_1	V_2	V_3
A	108	21.23	8.66
B	89	23.56	8.92
C	90	21.56	8.76
D	112	21.96	8.62

in-depth discussion of PCA, e.g. see Denis, 2021). In exploratory analysis, PCA has only a descriptive purpose. In particular, if the variance explained by the first component is high, most of the indicators correlate and represent a single aspect of the phenomenon. This leads to the conclusion that we can consider only one latent factor and then construct a single composite. Otherwise, if the variance explained by the first component is not very high, there are several groups of indicators representing different aspects of the phenomenon, and consequently, this seems to highlight the presence of more than one latent factor and the necessity of constructing more than one composite. There is no precise threshold; in general, if the first component explains more than 50% of the total variance, we can consider only one latent construct present (Alaimo & Maggino, 2020). The absence of correlation among the components is an useful property because it implies that the principal components measure different statistical dimensions in the data. It must be noted that PCA does not always work in the sense that a large number of original variables are reduced to a small number of transformed variables. Indeed, if the original variables are uncorrelated, the analysis does nothing. The best results were obtained when the original variables were highly correlated, positive, or negative. This is a crucial finding. The first principal component, resulting from PCA, is often used as a composite indicator. However, it represents highly intercorrelated indicators and neglects others. Therefore, many highly important but poorly intercorrelated indicators may not be represented by the composite index. In a formative model, this is not a good strategy because an indicator not correlated with the others measures a different aspect of the phenomenon.

At this point, we focus on the technical steps of normalisation and aggregation. To facilitate their explanation, we used an example of a system of three indicators and four units, as reported in Table 2.

Normalisation is required to make indicators comparable because they often present different measurement units and ranges. The objective is to transform them into pure numbers. Given the original data matrix \mathbf{X} , the objective is to obtain a matrix $\mathbf{R} \equiv \{r_{ij}\}$ where r_{ij} is the normalised value of the j -th indicator for the i -th unit. Normalisation is a very delicate step because it can change the distribution and the internal variability of the indicators. There are various normalisation methods. We report some of the most common normalisation methods, each of which has advantages and disadvantages. Choosing one rather than another affects synthesis. This problem can be partially overcome by performing a robustness analysis to evaluate the effects of the different procedures on the results obtained.

However, from a conceptual point of view, normalisation does not solve the problem of combining different measures, of *mixing apples and oranges* (Alaimo, 2022).

In normalisation, it is necessary to define the polarity of the basic indicators, that is, the sign of the relationship between the indicator itself and the phenomenon. Therefore, the type of composite we want to construct defines the polarity. In other words, some indicators may be positively related to the phenomenon to be measured (positive polarity), whereas others may be negatively related (negative polarity). For instance, if we want to construct a composite whose increase coincides with an improvement in well-being, job satisfaction would have a positive polarity, while the unemployment rate would be negative. On the contrary, if we want to construct a composite whose increase indicates a worsening of well-being, job satisfaction would have negative polarity, while the unemployment rate would be positive. After normalisation, all indicators must have positive polarity, that is, an increase in the normalised indicators corresponds to an increase in the composite index (Maggino, 2017: 166). If some indicators have a negative polarity, they must be inverted. There are two main methods for inverting polarity:

1. The linear transformation involves taking the complement with respect to the maximum value. Given the original data matrix \mathbf{X} , it is calculated as follows:

$$x'_{ij} = \max_i (x_{ij}) - x_{ij}$$

where x_{ij} is the value of the j -th indicator in the i -th unit, $\max_i (x_{ij})$ is the maximum value of the j -th indicator, and x'_{ij} is the inverted value. This is the simplest technique, which allows us to save the same distances between units with different origins. It is particularly used with ranking, standardisation, and rescaling normalisation methods.

2. The non-linear transformation consists of taking the reciprocal of the value. Given the original data matrix \mathbf{X} , it is calculated as follows:

$$x'_{ij} = \frac{1}{x_{ij}}$$

where x_{ij} is the value of the j -th indicator in the i -th unit, and x'_{ij} is the inverted value. This technique, typically used with indicisation, has been criticised because it modifies the distances between units and requires all values greater than 0.

Table 3 reports the results of the two inversion procedures for indicator V_3 .

A particular situation is the so-called double polarity, in which we observe an indicator presenting a positive polarity below a certain threshold and a negative polarity above it, or vice versa. Examples of such indicators are female-to-male ratios, that is, the ratio between the percentage of females and the percentage of males. These indicators are particularly used for measuring the gender gap (WEF,

Table 3 Example: System of three cardinal indicators observed in four units; linear and non-linear inversion of polarity for indicator V_3

	V_1	V_2	V_3	V'_3 (linear)	V'_3 (non-linear)
A	108	21.23	8.66	0.26	0.115
B	89	23.56	8.92	0.00	0.112
C	90	21.56	8.76	0.16	0.114
D	112	21.96	8.62	0.30	0.116

Table 4 Example: System of three cardinal indicators observed in four units: ranking normalisation

	V_1	V_2	V_3
A	3	1	2
B	1	4	4
C	2	2	3
D	4	3	1

2021): they have a positive polarity up to the value of 1 (which expresses gender equality between women and men); from 1 on, the polarity is reversed. Dealing with double polarity, we can use the triangular transformation

$$x'_{ij} = |\lambda_{x_j} - x_{ij}|$$

where x_{ij} is the value of the j -th indicator in the i -th unit, x'_{ij} is the inverted value, and λ_{x_j} is the value of the j -th indicator in which the polarity inverts (the threshold).

If all the indicators present the same unit of measurement and similar ranges or are expressed as percentages or ratios, a good choice is no normalisation, that is, aggregating the data of the original matrix. However, in most cases, we do not deal with such a situation; hence, we need to normalise.

Ranking

The normalised values of the j -th indicator are obtained by ranking its values in all statistical units:

$$r_{ij} = rank(x_{ij})$$

Thus, r_{ij} is the rank of the i -th unit in the ranking corresponding to the j -th indicator. If two or more units have the same value, several procedures can be used to assign a rank. One of the most widely used methods consists of assigning the same rank equal to the mean of the ranks they would have had in the case of different values. The transformation to ranks purifies indicators from the measurement unit. Its main advantage is that it is unaffected by the presence of outliers in the original data. However, ranking assumes the same distance between every unit, and consequently, the differences between units cannot be evaluated because absolute level information is lost. In Table 4, we report the results of ranking normalisation for the reported example.

Table 5 Example: system of three cardinal indicators observed in four units; min–max normalisation

	V_1	V_2	V_3
A	0.83	0.00	0.14
B	0.00	1.00	1.00
C	0.04	0.14	0.47
D	1.00	0.31	0.00

Re-scaling or Min–Max

The normalised values of the j -th indicator were re-scaled in the range $[0, 1]$ as follows:

$$r_{ij} = \frac{x_{ij} - \min_i(x_{ij})}{\max_i(x_{ij}) - \min_i(x_{ij})}$$

where $\max_i(x_{ij})$ and $\min_i(x_{ij})$ are, respectively, the minimum and maximum values (commonly the observed values in the N statistical units) that represent the possible range of the j -th indicator. Reporting an indicator in the range $[0, 1]$ can be an advantage, giving an easy-to-read representation. Moreover, the range of indicators with very little variation will increase, which will contribute more to the composite (this is evident in the example in Table 5). The main drawback is that being based on the range, it is sensitive to outliers. In Table 5, we report the results of the min–max normalisation for the reported example.

Standardisation or z-scores

The normalised values of the j -th indicator were obtained as z-scores, converting the indicator to a common scale with 0 mean and standard deviation equal to 1, as follows:

$$r_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}$$

where $\mu_j = \frac{\sum_{i=1}^N x_{ij}}{N}$ and $\sigma_j = \sqrt{\frac{\sum_{i=1}^N (x_{ij} - \mu_j)^2}{N}}$ are the arithmetic mean and standard deviation of the indicator j -th. The main advantage of this method is that it reports the indicator to a standard Gaussian distribution and, consequently, simplifies the analysis. The main drawback is the presence of negative values, which can be a limitation of some aggregation methods (i.e. geometric mean). In Table 6, we report the results of the z-score normalisation for the reported example.

Indicisation

The normalised values of the j -th indicator are obtained as percentage ratios between the original values and a reference, as follows:

Table 6 Example: System of three cardinal indicators observed in four units: z-scores normalisation

	V ₁	V ₂	V ₃
A	0.399	-0.822	-0.597
B	-1.214	1.436	1.343
C	-1.129	-0.502	0.149
D	0.739	-0.114	-0.896
μ_j^a	103.300	22.078	8.740
σ_j^a	11.776	1.032	0.134

^aArithmetic mean and standard deviations of indicators are calculated based on the original values reported in Table 2

Table 7 Example: System of three cardinal indicators observed in four units: indicisation

	V ₁	V ₂	V ₃
A	96.429	90.110	97.085
B	79.464	100.000	100.000
C	80.357	91.511	98.206
D	100.000	93.209	96.937

$$r_{ij} = \frac{x_{ij}}{x_{oj}} * 100$$

where x_{oj} is the reference value selected for the j -th indicator, which generally corresponds to the maximum observed or to a general benchmark. This method makes it possible to decouple indicators from the unit of measurement and to preserve the relative distance between different units. The main drawback of this method is its high sensitivity to outliers. In Table 7, we report the result of the indicisation for the example reported using the maximum value observed in each indicator as a reference value.

The following step is the aggregation of normalised indicators, that is, the composition of the normalised indicators into a single synthetic index. In the literature, many methods have been proposed for constructing composites (there is no objective of this chapter to report a review of all aggregation methods and procedures in the literature; for more detailed information, see Saisana & Tarantola (2002), OECD (2008), and Maggino (2017)). Each method has its advantages and disadvantages; there is no such thing as the best method. The method used has an impact on the results obtained, in particular, the definition of the importance of each individual indicator (weighting) and the identification of the technique for synthesising the indicators.

The choice of weighting has a large impact on the values and consequently on the meaning of the composites. Thus, it is essential to understand the effects of one choice on another. In the literature, there are different approaches to the weighting issue, which can be traced to three categories (Gan et al. 2017):

- Giving to all the indicators the same weight (equal weighting)
- Weights derived from the statistical characteristics of the data and attributed as the result of a statistical method, for instance, principal component analysis (statistic-based weighting)
- Weights assigned to individual indicators based on the judgments of the public or experts (public/expert opinion-based weighting)

No agreed-upon methodology exists to weigh basic indicators. The simplest weighting strategy, that is, attributing equal weight to all basic indicators, considering them equally important (Nardo et al., 2005) is the most commonly used. This method is not without criticism, especially from those who consider a possible misconception of the underlying logic according to which the weight assigned to a variable can be directly interpreted as a measure of its importance to the value of the composite (Becker et al., 2017: 12). The statistical method, for instance, using the results of PCA, is very questionable because most of the time it is based on the correlations among basic indicators and, as we have seen, their interpretation changes according to the measurement model. It is likely that the best method is based on the opinions of stakeholders and experts. When the latter cannot be used, a good strategy could be to select a limited number of robust indicators, giving them the same weight (Alaimo, 2022).

Aggregation methods can be classified according to various criteria (Gan et al., 2017). One of the main classifications is based on the degree of toleration/substitutability among the basic indicators. The components of a synthetic index are called substitutable if a deficit in one component can be compensated for by a surplus in another. The assumption of component substitutability implies the adoption of additive aggregation methods (e.g. arithmetic mean). The components are defined as nonsubstitutable if no compensation is allowed between them. In this case, multiplicative (e.g. geometric mean) or noncompensative methods are adopted. Thus, this conceptual assumption has an important effect on the other steps of the construction of the composites, in particular, the selection of the aggregation function. Based on this classification criterion, we can distinguish between the following:

- *Additive aggregation methods*: They employ functions that sum the normalised values of the basic indicators to form a composite index. The most widely used additive method is the *weighted arithmetic mean*. Given the normalised matrix $\mathbf{R} \equiv \{r_{ij}\}$, the value of the composite indicator C_i for generic unit i -th is obtained as follows:

$$C_i = \frac{\sum_{j=1}^M r_{ij} w_j}{M}$$

where w_j is the weight of the j -th indicator. The weights must satisfy the following constraints: $w_j > 0$ and $\sum_{j=1}^M w_j = 1$. In the case of equal weighting, that is, $w_j = \frac{1}{M}$, we have the simple arithmetic mean. This technique implies full compensability such

Table 8 Example: System of three cardinal indicators observed in four units; min–max normalisation; arithmetic and geometric mean aggregation

	V_1	V_2	V_3	C_i (arithmetic)	C_i (geometric)
A	0.83	0.00	0.14	0.323	0.000
B	0.00	1.00	1.00	0.667	0.000
C	0.04	0.14	0.47	0.217	0.138
D	1.00	0.31	0.00	0.437	0.00

that poor performance in some indicators can be compensated for by sufficiently high values in other indicators.

- *Multiplicative aggregation methods:* Multiplicative functions are used on the normalised values of basic indicators to form a composite index. The most widespread method is the *weighted geometric mean*. Given the normalised matrix $\mathbf{R} \equiv \{r_{ij}\}$, the value of the composite indicator C_i for generic unit i -th is obtained as follows:

$$C_i = \sqrt[M]{\prod_{j=1}^M r_{ij}^{w_j}}$$

where w_j is the weight of the j -th indicator. The weights must satisfy the following constraints: $w_j > 0$ and $\sum_{j=1}^M w_j = 1$. In the case of equal weighting, that is, $w_j = \frac{1}{M}$, we have a simple geometric mean. Geometric mean-based methods only allow compensability between indicators within certain limitations (partially compensative) because of the *geometric-arithmetic mean inequality* (Beliakov et al., 2007), which limits the ability of indicators with very low scores to be fully compensated for by indicators with high scores.

In Table 8, we report the results of the aggregation using simple arithmetic and geometric means for the values normalised with min-max (Table 5).

Additive and multiplicative methods imply total and partial compensation, respectively, among the basic indicators. The compensability issue is not only methodological but also, and above all, conceptual. Choosing one approach over the other affects not only the values of the composite but also, and more importantly, the interpretation of the phenomenon being measured. For instance, looking at the Human Development Index (UNDP, 1990), if we admit full compensability, we implicitly affirm that a surplus in education can compensate for a deficit in health. This is highly questionable. However, if we affirm the non-compensability of the basic indicators, we risk crushing the results of our synthesis. A possible solution identified in the literature (Casadio Tarabusi & Guarini, 2013; Mazziotta & Pareto, 2016) is the adoption of a partially compensative method, that is allowing compensation ‘up to a certain point’; however, the question would arise as to what is the permissible and tolerable threshold of compensability.

The Benefit of the Doubt (BoD) approach is an aggregative method for composite indicator construction (Cherchye et al., 2007; Rogge, 2018) based on Data

Envelopment Analysis (DEA), a linear programming technique that is useful for measuring the relative efficiency of decision-making units on the basis of multiple inputs and outputs (Farrell, 1957; Charnes et al., 1978). The efficiency of a set of indicators can be adapted to construct a synthetic indicator using input-oriented DEA. The synthetic measure is obtained as the weighted sum of the normalised indicators relative to a benchmark. More precisely, it is defined as the performance of a single unit divided by the performance of the benchmark:

$$\text{BoD}_i = \frac{\sum_{j=1}^M r_{ij} w_{ij}}{r_{ij}^*}$$

where r_{ij} is the normalised value of the j -th indicator for the i -th statistical unit according to the min–max procedure, w_{ij} is the corresponding weight, and r_{ij}^* is the benchmark given by the following:

$$r_{ij}^* = \max_{i \in [1, \dots, N]} \sum_{j=1}^M r_{ij} w_{ij}$$

The identification of the optimal set of weights guarantees that each unit is associated with the best possible position compared to all the others. Optimal weights were obtained by solving the following equation:

$$\text{BoD}_i^* = \max_{w_{ij}} \frac{\sum_{j=1}^M r_{ij} w_{ij}}{\max_{k \in [1, \dots, N]} \sum_{j=1}^M r_{kj} w_{kj}}, \forall i = 1, \dots, N$$

under the constraint that the weights are non-negative, and the result is bounded [0, 1]. The most favourable weights were always applied to all observations. The main advantages of this method are related to the DEA solution. Because the weights are specific for each unit, cross-unit comparisons are not possible, and the values of the scoreboard depend on the benchmark performance. Another drawback is the multiplicity of the equilibria. Hiding the problem of multiple equilibria prevents the weights from being uniquely determined (even if the composite indicator is unique). The optimisation process could lead to many 0-weights if no restrictions were imposed on the weights.

The construction of a composite involves different subjective choices: the selection of individual indicators, choice of aggregation model, and weights of the indicators. All these subjective choices are the bones of the composite indicator, and together with the information provided by the numbers themselves, shape the message communicated by the composite indicator (OECD, 2008). The effectiveness of a composite index also depends on testing its assumptions, which is the purpose of the validation. It evaluates the robustness of the composite index in terms of its capacity to produce correct and stable measures and its discriminant capacity (Maggino, 2017). The robustness of a composite index is assessed by uncertainty

analysis, which focuses on how uncertainty in the input factors propagates through the structure of the composite index and affects the results. The sensitivity analysis focuses on how much each individual source of uncertainty contributes to the output variance (Saisana et al., 2005). Used during composite construction, these procedures help in indicator selection, add transparency to the index construction process, and explore the robustness of alternative composite index designs and rankings. The discriminant capacity of a composite index is assessed by exploring its capacity to discriminate between units and/or groups, distributing all the units without any concentration of individual scores in a few segments of the continuum, showing values that are interpretable in terms of selectivity through the identification of particular reference values or cut-off points (Maggino & Zumbo, 2011).

References

- Alaimo, L. S. (2022). *Complexity of social phenomena: Measurements, analysis, representations and synthesis*. Sapienza University Press.
- Alaimo, L. S. (2021a). Complex systems and complex adaptive systems. In F. Maggino (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 1–3). Springer. https://doi.org/10.1007/978-3-319-69909-7_104659-1
- Alaimo, L. S. (2021b). Complexity and knowledge. In F. Maggino (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 1–2). Springer. https://doi.org/10.1007/978-3-319-69909-7_104658-1
- Alaimo, L. S., Arcagni, A., Fattore, M., & Maggino, F. (2021a). Synthesis of multi-indicator system over time: A poset-based approach. *Social Indicators Research*, *157*(1) 77–99.
- Alaimo, L. S., Ivaldi, E., Landi, S., & Maggino, F. (2021b). Measuring and evaluating socioeconomic inequality in small areas: An application to the urban units of the municipality of Genoa. *Socio-Economic Planning Sciences*, *83*, 101170.
- Alaimo, L. S., Arcagni, A., Fattore, M., Maggino, F., & Quondamstefano, V. (2022a). Measuring equitable and sustainable well-being in Italian regions. The non-aggregative approach. *Social Indicators Research*, *161*(2), 711–733.
- Alaimo, L. S., Fiore, M., & Galati, A. (2022b). Measuring consumers' level of satisfaction for online food shopping during COVID-19 in Italy using posets. *Socio-Economic Planning Sciences*, *82*, 101064.
- Alaimo, L. S., Galli, E., Rizzo, I., & Scaglioni, C. (2023). A new index of transparency: Evidence for the Italian municipalities. *Socio-Economic Planning Sciences*, *89*, 101678. <https://doi.org/10.1016/j.seps.2023.101678>
- Alaimo, S. L., & Maggino, F. (2020). Sustainable development goals indicators at territorial level: Conceptual and methodological issues—The Italian perspective. *Social Indicators Research*, *147*(2), 383–419.
- Arrow, K. J. (2012). *Social choice and individual values*. Yale University Press.
- Becker, W., Saisana, M., Paruolo, P., & Vandecasteele, I. (2017). Weights and importance in composite indicators: Closing the gap. *Ecological Indicators*, *80*, 12–22.
- Beliakov, G., Pradera, A., & Calvo, T. (2007). *Aggregation functions: A guide for practitioners* (Vol. 221). Springer.
- Belton, V., & Stewart, T. (2002). *Multiple criteria decision analysis—An integrated approach*. Kluwer Academic Press.
- Bertalanffy, L. (1968). *General system theory: Foundations, development, applications*. George Braziller.

- Blalock, H. M. (1964). *Causal inferences in nonexperimental research*. University of North Carolina Press.
- Blalock, H. M. (1968). The measurement problem: A gap between the languages of theory and research. In F. Kerlinger (Ed.), *Methodology in social research* (pp. 5–27). McGrawHill.
- Bollen, K. A. (1984). Multiple indicators: Internal consistency or no necessary relationship? *Quality and Quantity*, 18(4), 377–385.
- Bollen, K. A. (2007). Interpretational confounding is due to misspecification, not to type of indicator: Comment on Howell. *Psychological Methods*, 12(2), 219–228.
- Bollen, K. A., & Diamantopoulos, A. (2017). In defense of causal-formative indicators: A minority report. *Psychological Methods*, 22(3), 581–596.
- Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110(2).
- Bruggemann, R. R., & Patil, G. P. (2011). *Ranking and prioritization for multi-indicator systems: Introduction to partial order applications*. Springer Science & Business.
- Capra, F. (1996). *The Web of Life: A New Scientific Understanding of Living Things*. Anchor books.
- Cattell, R. B. (1978). *The use of factor analysis in behavioral and life sciences*. Plenum.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research*, 2(6), 429–444.
- Cherchye, L., Moesen, W., Rogge, N., & Puyenbroeck, T. (2007). An introduction to ‘benefit of the doubt composite indicators’. *Social Indicators Research*, 82(1), 111–145.
- Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research*, 61(12), 1250–1262.
- Curtis, R. F., & Jackson, E. F. (1962). Multiple indicators in survey research. *American Journal of Sociology*, 68(2), 195–204.
- D’Urso, P. (2000). Dissimilarity measures for time trajectories. *Statistical Methods & Applications*, 9(1–3), 53–83.
- Denis, D. J. (2021). *Applied univariate, bivariate, and multivariate statistics: Understanding statistics for social and natural scientists, with applications in SPSS and R*. Wiley.
- Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research*, 61(12), 1203–1218.
- Diamantopoulos, A., & Siguaw, J. A. (2006). Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. *British Journal of Management*, 17(4), 263–282.
- Diener, E., & Suh, E. (1997). Measuring quality of life: Economic, social, and subjective indicators. *Social Indicators Research*, 40(1), 189–216.
- Edwards, J. R. (2011). The fallacy of formative measurement. *Organizational Research Methods*, 14(2), 370–388.
- Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, 5(2), 155–174.
- Ehrgott, M., Figueira, J., & Greco, S. (Eds.). (2005). *Multiple criteria decision analysis: State of the art surveys*. Springer.
- Érdi, P. (2008). *Complexity explained*. Springer.
- Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society: Series A (General)*, 120(3), 253–281.
- Fattore, M. (2016). Partially ordered sets and the measurement of multidimensional ordinal deprivation. *Social Indicators Research*, 128(2), 835–858.
- Fattore, M. & Alaimo, L.S. (2023) A partial order toolbox for building synthetic indicators of sustainability with ordinal data. *Socio-Economic Planning Sciences*, 88, 101623.
- Gan, X., Fernandez, I. C., Guo, J., Wilson, M., Zhao, Y., Zhou, B., & Wu, J. (2017). When to use what: Methods for weighting and aggregating sustainability indicators. *Ecological Indicators*, 81, 491–502.
- Haken, H. (1983). *Synergetics: An introduction*. Springer.

- Hennig, C., Meila, M., Murtagh, F., Fionn, & Rocci, R. (2015). *Handbook of cluster analysis*. CRC Press.
- Holland, J. H. (1992). Complex adaptive systems. *Daedalus*, 121(1), 17–30.
- Horn, R. V. (1993). *Statistical indicators: For the economic and social sciences*. Cambridge University Press.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6), 417–441.
- Howell, R. D., Breivik, E., & Wilcox, J. B. (2007). Reconsidering formative measurement. *Psychological Methods*, 12(2), 205–218.
- Kerber, A. (2017). Evaluation, considered as problem orientable mathematics over lattices. In M. Fattore & R. Brüggemann (Eds.), *Partial order concepts in applied sciences* (pp. 87–103). Springer.
- Land, K. C. (1970). On the estimation of path coefficients for unmeasured variables from correlations among observed variables. *Social Forces*, 48(4), 506–511.
- Landau, S., Leese, M., Stahl, D., Daniel, & Everitt, B. S. (2011). *Cluster analysis*. Wiley.
- LaZarsfeld, P. F. (1958). Evidence and inference in social research. *Daedalus*, 87(4), 99–130.
- Letiche, H., Lissack, M., & Schultz, R. (2012). *Coherence in the midst of complexity: Advances in social complexity theory*. Palgrave Macmillan.
- Lima, M. (2013). *Visual complexity: Mapping patterns of information*. Princeton Architectural Press.
- Macoun, P., & Prabhu, R. (1999). *Guidelines for applying multi-criteria analysis to the assessment of criteria and indicators* (Vol. 9). CIFOR.
- Maggino, F. (Ed.). (2017). *Complexity in society: From indicators construction to their synthesis*. Springer.
- Maggino, F., & Alaimo, L. S. (2021). Complexity and wellbeing: Measurement and analysis. In L. Bruni, A. Smerilli, & D. Rosa (Eds.), *A modern guide to the economics of happiness* (pp. 113–128). Edward Elgar.
- Maggino, F., & Alaimo, L. S. (2022). Measuring complex socioeconomic phenomena. Conceptual and methodological issues. In S. Valaguzza & M. A. Hughes (Eds.), *Interdisciplinary approaches to climate change for sustainable growth* (pp. 43–59). Springer.
- Maggino, F., Brüggemann, R., & Alaimo, L. S. (2021). Indicators in the framework of partial order. In R. Brüggemann, L. Carlsen, T. Beycan, C. Suter, & F. Maggino (Eds.), *Measuring and understanding complex phenomena* (pp. 17–29). Springer.
- Maggino, F., & Zumbo, B. (2011). Measuring the quality of life and the construction of social indicators. In K. C. Land, A. C. Michalos, & J. M. Sirgy (Eds.), *Handbook of social indicators and quality of life research* (pp. 201–238). Springer.
- Maharaj, E. A., D’Urso, P., & Caiado, J. (2019). *Time series clustering and classification*. CRC Press.
- Mazziotta, M., & Pareto, A. (2016). On a generalized noncompensatory composite index for measuring socioeconomic phenomena. *Social Indicators Research*, 127(3), 983–1003.
- McLean, I. (1990). The Borda and Condorcet principles: Three medieval applications. *Social Choice and Welfare*, 7(2), 99–108.
- Meadows, D. H. (2009). *Thinking in systems: A primer*. Earthscan.
- Michell, J. (1999). *Measurement in psychology: A critical history of a methodological concept*. Cambridge University Press.
- Morin, E. (1977). *La méthode. Tome I. La nature de la nature*. Le Seuil.
- Morin, E. (1985). Le vie della complessità. In G. Bocchi & M. Ceruti (Eds.), *La sfida della complessità* (pp. 49–60). Feltrinelli.
- Nardo, M., Saisana, M., Saltelli, A., & StefanoTarantola. (2005). *Tools for composite indicators building*. European Commission. Institute for the Protection and Security of the Citizen.
- Neggers, J., & Kim, H. S. (1998). *Basic posets*. World Scientific Publishing.
- Nijkamp, P., & Delft, A. (1977). *Multi-criteria analysis and regional decision-making* (Vol. 8). Springer Science & Business Media.

- O.E.C.D. (2008). *Handbook on constructing composite indicators. Methodology and user guide*. OECD Publications.
- Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(11), 559–572.
- Rogge, N. (2018). On aggregating benefit of the doubt composite indicators. *European Journal of Operational Research*, 264(1), 364–369.
- Russell, B. (2009). *Principles of mathematics*. Routledge.
- Sacconaghi, R. (2017). Building knowledge. Between measure and meaning: A phenomenological approach. In F. Maggino (Ed.), *Complexity in society: From indicators construction to their synthesis* (pp. 51–68). Springer.
- Saisana, M., & Tarantola, S. (2002). State-of-the-art report on current methodologies and practices for composite indicator development. In *EUR 20408 EN*.
- Saisana, M., Tarantola, S., & Saltelli, A. (2005). Uncertainty and sensitivity techniques as tools for the analysis and validation of composite indicators. *Journal of the Royal Statistical Society A*, 168, 307–323.
- Salzman, J. (2003). *Methodological choices encountered in the construction of composite indices of economic and social well-being*. (Technical Report). Center for the Study of Living Standards.
- Schroder, B. (2002). *Ordered set. An introduction*. Birkauer.
- Sen, A. (1977). Social choice theory: A re-examination. *Econometrica: Journal of the Econometric Society*, 53–89.
- Spearman, C. (1904). General intelligence objectively determined and measured. *American Journal of Psychology*, 15(2), 201–293.
- Stengers, I. (1985). Perché non può esserci un paradigma della complessità. In G. Bocchi & M. Ceruti (Eds.), *La sfida della complessità* (pp. 61–83). Feltrinelli.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684), 677–680.
- Tarabusi, E. C., & Guarini, G. (2013). An unbalance adjustment method for development indicators. *Social Indicators Research*, 112(1), 19–45.
- Thurstone, L. (1931). Multiple factor analysis. *Psychological Review*, 38(5), 406–427.
- Toni, A. F., & Comello, L. (2010). *Journey into complexity*. Marsilio Editori.
- Tufte, E. R. (2001). *The visual display of quantitative information* (Vol. 2). Graphics Press.
- U.N.D.P. (1990). *Human development report 1990*. Oxford University Press.
- W.E.F. (2021). *Global gender gap report 2021*. World Economic Forum.
- Waldrop, M. M. (1992). *Complexity: The emerging science at the edge of order and chaos*. Simon and Schuster.
- Wilcox, J. B., Howell, R. D., & Breivik, E. (2008). Questions about formative measurement. *Journal of Business Research*, 61(12), 1219–1228.
- Zopounidis, C., & Pardalos, P. M. (2010). *Handbook of multicriteria analysis* (Vol. 103). Springer Science & Business Media.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

