Palacký University Olomouc

University of Clermont Auvergne

University of Pavia

MASTER THESIS

Transparency of Composite Indicators of Development

Fabiola Cañas Magaña

Supervisor: Professor Miroslav Syrovátka

GLODEP 2021

Palacký University Olomouc
University of Clermont Auvergne
University of Pavia

Transparency of Composite Indicators of Development

Fabiola Cañas Magaña
Supervisor: Professor Miroslav Syrovátka

GLODEP 2021

Declaration

I, Fabiola Cañas Magaña, hereby declare that the Master Thesis "Transparency of Composite Indicators of Development", submitted to the Erasmus Mundus Joint Master Degree in International Development Studies GLODEP Consortium, is my original work, except where it is explicitly stated by referencing other sources. This document has not been submitted to any other organization or university.

May 2021

UNIVERZITA PALACKÉHO V OLOMOUCI

Přírodovědecká fakulta

Akademický rok: 2020/2021

ZADÁNÍ DIPLOMOVÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení:	**Lic. Fabiola CANAS MAGANA**
Osobní číslo:	**R190706**
Studijní program:	**N1301 Geography**
Studijní obor:	**International Development Studies**
Téma práce:	**Transparency of Composite Indicators of Development**
Zadávající katedra:	**Katedra rozvojových a environmentálních studií**

Zásady pro vypracování

Development and related concepts, such as well-being, are inherently multidimensional. This renders attempts to measure them a difficult task, as there is no direct way to assess them. Composite indicators have become increasingly popular due to their ability to comprise complex concepts in a single measure. However, significant criticism has also been risen since this type of indicators are created through a series of assumptions and decisions made by their authors, leaving significant scope for arbitrariness in their construction. Credible construction of composite indicators is therefore essential and should assume transparency in all methodological procedures, including their justification.
This research will explore and analyze composite indicators in terms of transparency, by reviewing general recommendations for their construction and critical views on the methodology of existing indicators. The aim is to provide recommendations on the construction and dissemination of composite indicators of development.

Rozsah pracovní zprávy:	**20-25 tisíc slov**
Rozsah grafických prací:	**dle potřeby**
Forma zpracování diplomové práce:	**tištěná**
Jazyk zpracování:	**Angličtina**

Seznam doporučené literatury:

Barrington-Leigh, C., Escande, A. 2018. Measuring Progress and Well-Being: A Comparative Review of Indicators. *Social Indicators Research* 135 (3), pp. 893–925.
Becker, W., Saisana, M., Paruolo, P., Vandecasteele, I. 2017. Weights and importance in composite indicators: Closing the gap. *Ecological Indicators* 80, pp. 12–22.
Freudenberg, M. 2003. Composite Indicators of Country Performance: A Critical Assessment. *OECD Science, Technology and Industry Working Papers*, No. 2003/16. OECD Publishing.
Greco, S., Ishizaka, A., Tasiou, M., Torrisi, G. 2019. On the Methodological Framework of Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness. *Social Indicators Research* 141 (1), pp. 61–94.
Lustig, N. 2011. Multidimensional indices of achievements and poverty: what do we gain and what do we lose? An introduction to JOEI Forum on multidimensional poverty. *Journal of Economic Inequality* 9 (2), pp. 227–234.
OECD. 2008. *Handbook on Constructing Composite Indicators. Methodology and User Guide*. Paris: OECD.
Ravallion, M. 2012. Mashup Indices of Development. T*he World Bank Research Observer* 27 (1), pp. 1–32.

Vedoucí diplomové práce: **Mgr. Miroslav Syrovátka, Ph.D.**
Katedra rozvojových a environmentálních studií

Datum zadání diplomové práce: **29. ledna 2021**
Termín odevzdání diplomové práce: **31. května 2021**

L.S.

doc. RNDr. Martin Kubala, Ph.D.
děkan

doc. RNDr. Pavel Nováček, CSc.
vedoucí katedry

V Olomouci dne 29. ledna 2021

Acknowledgments

I would like to express my sincerest gratitude to my Supervisor, Professor Miroslav Syrovátka, for his time, guidance, and the many valuable comments he provided, which have significatively improved this project.

To my GLODEP 3rd cohort friends, I give a heartfelt acknowledgment, as we have supported each other not only academically but in every other aspect as well. In them, I have found a family while I was far from home. I will always be thankful for that.

I would like to thank everyone that has helped me in this process. The constant support and help I have received from my family, friends, and professors have been crucial to finish this program. I especially thank Mami, Papi, Toti, and Pa for always believing in me.

Abstract

Development is a multidimensional concept that has been tried to be captured through various types of measures, one of them being composite indicators. These indicators are praised for their ability to comprise complex concepts into a single number; however, they face significant criticism since they are created through a series of subjective decisions, leaving significant scope for arbitrariness. To present a credible measure, authors should procure a transparent disclosure of the construction process and its results, thus contributing to avoid misinterpretation and misuse of the proposed measure. Therefore, the present study explores the transparency with which the construction process of three composite indicators of development is disclosed, with the purpose of highlighting aspects that require improvement. The studied composite indicators are the Social Progress Index, the Human Development Index, and the Legatum Prosperity Index. For the most recent version of the methodology of each index, the analysis finds that authors disclose their methodological decisions, though justifications, discussions of alternatives, and the implications of such decisions are often omitted. Therefore, the study advocates for improving the transparency in communicating the methodology and properties of each index, and to explain them in adherence to the adopted theoretical framework.

Keywords: development, composite indicator, transparency, human development, prosperity, social progress.

Contents

List of tables

List of figures

List of abbreviations

DTF	Distance To Frontier
GDP	Gross Domestic Product
GNI	Gross National Income
HDI	Human Development Index
HDR	Human Development Report
HDRO	Human Development Report Office
JRC	European Union Joint Research Centre
LPI	Legatum Prosperity Index
MRS	Marginal Rate of Substitution
OECD	Organization for Economic Cooperation and Development
PCA	Principal Component Analysis
SPI	Social Progress Index
UNDP	United Nations Development Programme
UNECE	United Nations Economic Comission for Europe

Introduction

Development measurement is a field that gained momentum starting from the last decades of the 20th century. Nowadays, there is a consensus that economic indicators can only capture a part of a country's situation in terms of living standard and well-being of its population. This transition meant going beyond traditional measures of output, such as the Gross Domestic Product (GDP), and beginning a quest for a measure that could capture a complex concept, whose definition is -to this day- a reason for debate; non surprisingly, the way to measure development, and its various dimensions, is even further from a widespread consensus.

Nevertheless, various advancements and milestones can already be traced: the issuing of the first Human Development Report (HDR) in 1990 -and with it the first Human Development Index (HDI)-; the publication in 1999 of Amartya Sen's seminal book: *Development as Freedom*; the Beyond GDP initiative high-level conferences held in Europe in 2007; the creation of "The Commission on the Measurement of Economic Performance and Social Progress" – that culminated in the publication of the well-known Stiglitz-Sen-Fitoussi report; among others.

Different types of measures that try to capture development and related concepts have been proposed in the past decades. Barrington-Leigh and Escande (2018) conducted a review of progress and well-being measures published since the 1970s. Among them, one can identify at least two broad categories: aggregate indicators, which, through various methodological processes, produce synthetic measures, and sets of indicators (dashboards), that is, collections of indicators presented together to portray a complex picture. The focus of the present study is on the former, specifically on composite indicators of development.

Composite indicators have been subject to constant criticism, which predominantly stems from the absence of a theory that guides the decisions that their creation process entails. In fact, the inherent subjectivity is considered by some scholars as the core argument wielded against composites (Saltelli, 2007, p. 69). When the composite intends to measure a socioeconomic phenomenon, such as development and related concepts, the criticism goes beyond the purely methodological arena since the reasoning behind its creation implies normative assumptions. The creation of the index gives a clear indication of *what* the authors consider that drives a positive or a negative change and, depending on the chosen indicators, some idea of *how* it does it.

However, composite indicators' ability to comprise complex concepts into a single measure has many advantages, ranging from facilitating communication to the public to contributing to policy analysis and decision making. These are some of the reasons why, despite the criticism, composite indicators are continually being created. In fact, such criticism has in many cases propelled corrections and even serve as inspiration for new alternative composites to arise

For example, one can observe the abundant criticism surrounding arguably the most popular composite, the Human Development Index (HDI) published by the United Nations Development Programme (UNDP). Fruitfully such criticism has prompted a series of modifications to its methodology and led to the utilization of complementary composite indicators (such as the Gender Development Index) to provide a broader picture of the local -and global- development situation.

It is not likely that a consensus will be reached on the optimal way to create a composite indicator. Therefore, a minimum requirement is to be transparent in all the methodological choices, including their justification. Freudenberg (2003) states that indexes should always be accompanied by explanations of their components, construction, and interpretations, without omitting their weaknesses. Therefore, the present study intends to explore the transparency with which the methodology of selected composite indicators of development is presented.

Rather than an evaluation of the composite indicator's empirical soundness or theoretical adequacy, the present study highlights the importance of clear communication of the construction process of these aggregate development measures, with a focus on transparent dissemination of its construction process.

The following Chapter presents the methodology followed in this study. Chapter 2 presents the framework with which the selected composite indicators will be evaluated. Chapter 3 develops an individual analysis of the transparency of three composite indicators: the Social Progress Index (SPI), Human Development Index (HDI) and Legatum Prosperity Index (LPI). Subsequently, Chapter 4 presents a comparative discussion of the transparency of these composite indicators and highlights aspects that require improvement, including recommendations derived from the analysis. Lastly, the conclusions of the study are presented.

Chapter 1 Methodology for evaluating transparency of composite indicators of development

The present Chapter's objective is to explain the steps followed in the present study. Firstly, a framework for evaluating the transparency of composite indicators of development is developed (Chapter 2). This framework is based on general recommendations for constructing composite indicators and on experts' opinions critically discussing several of their methodological -and theoretical- aspects. The framework's structure is based on the sequential step guide developed by the Organization for Economic Cooperation and Development (OECD) (2008), the *Handbook on Constructing Composite Indicators*. The Handbook is widely recognized as a high standard set of applicable recommendations. However, disclosing evidence in this specific sequential structure or identifying every step without exclusion is not a requirement.

Secondly, a selection process of composite indicators of development was performed to identify those that would be evaluated. In order to map the included composite indicators, the surveys conducted by Bandura (2008), Yang (2014), and Barrington-Leigh and Escande (2018) were consulted. Three criteria were considered for selection:

a) The index tries to capture development in its multidimensionality: different approaches have been adopted for measuring development, while some of them focus on a specific dimension of it (i.e., gender, governance, among others), the focus of this study is on those that try to capture the concept in its multidimensionality.
b) Its scope is international: the composite indicator has been created to be applied internationally.
c) The index is currently being updated: the latest index results were recently published, and it is expected that it will be updated in the future.

The three composite indicators selected are depicted in Table 1.

Table 1 Selected composite indicators of development

Index	Measured concept	Year of creation	Number of countries ranked (2020)
Social Progress Index	Social progress	2013	163
Human Development Index	Human development	1990	189
Legatum Prosperity Index	Prosperity	2007	167

Source: author based on UNDP (2020a), Stern et al. (2020), and Legatum Institute (2020a).

This selection allows studying three well-established composite indicators, which have been published for several years and have benefited from expert feedback throughout their existence. For the three composite indicators, information about the theoretical framework and methodology was collected from the relevant websites of the Social Progress Imperative (socialprogress.org), the UNDP-HDR (hdr.undp.org) and the Legatum Institute (prosperity.com). Although the focus of the study is to evaluate the transparency with which the latest available methodology of the index is disclosed, reference to older version's documentation is necessary to provide a complete picture of the methodological design. The results of the analysis for each composite indicator are presented in Chapter 3.

Afterward, a comparative discussion of the transparency with which the three composite indicators are disclosed is presented in Chapter 4. The analysis follows the structure and evidence disclosed in Chapter 3. The objective is to highlight identified omissions -or just partial recognition- of important considerations about the construction process -and the index's properties- and to identify recommendations to improve the transparency of the methodology's disclosure. Due to the inherent theoretical and methodological heterogeneity between the three composite indicators, categorizing and ranking the level of transparency of the indexes is not the finality of this study.

Chapter 2 Framework for a transparency evaluation of composite indicators of development

The present Chapter describes the framework that is used to conduct the transparency evaluation. It determines what will be sought in the different types of resources used by authors to communicate the composite indicator's construction process and its results. As mentioned, the framework's structure is based on the steps in the construction of composite indicators proposed by the OECD (2008), which are the following[1]:

1. Theoretical framework
2. Data selection
3. Imputation of missing data
4. Multivariate analysis
5. Normalization
6. Weighting and aggregation
7. Uncertainty and sensitivity analysis
8. Back to the real data
9. Links to other indicators (or variables)

Given that the choices in each of the enlisted steps are closely related, four groups are proposed to conduct a comprehensive analysis of the evaluated composite indicators, as shown in Table 2. The essential criteria for evaluating transparency will be (1) that the authors disclose their decisions and (2) that they justify them. In the following pages, a description of what is expected to be disclosed regarding each step is provided.

An in-depth guide of practical information and methods for each step, including their associated advantages and disadvantages, can be found in OECD (2008), United Nations Economic Comission for Europe (UNECE) (2019), and other expert studies cited in this Chapter.

[1] The OECD (2008, p. 16) proposes as a tenth step the "presentation and visualization" of the composite results; however, the present study focuses on the transparency of the disclosure of the composite index's construction process, so the relevant aspects of the presentation of the data -and of the index's results- will be covered within the section to which they relate.

Table 2 Sections of the framework for a transparency evaluation of composite indicators of development

Section		Steps (OECD, 2008)
1.	Foundations of the composite indicator: theory and variables	➢ Theoretical framework ➢ Data selection ➢ Imputation of missing data
2.	Structural form of the index	➢ Multivariate analysis ➢ Normalization ➢ Weighting and aggregation
3.	Robustness tests	➢ Uncertainty and sensitivity analysis
4.	Analysis of the index results against its components and related indicators	➢ Back to the real data ➢ Links to other indicators (variables)

Source: author, based on OECD (2008).

2.1 Foundations of the composite indicator: theory and variables

2.1.1 Theoretical framework

The authors must clearly define *what* is being measured. The dimensions of development that are considered should equally be defined and justified, describing the relations between them both theoretically and empirically, when possible (OECD, 2008). Additionally, the authors may describe related phenomena and concepts (UNECE, 2019); this could contribute to expose the value added of the aggregate measure, by highlighting the similarities, differences, or relation it has with other indicators.

While defining the theoretical framework, authors should ideally also clearly define "the objectives and the end-users" of the index (Becker et al., 2019, p. 8). However, some authors find that many experts "often seem to produce indicators without a clear intended use or user in mind" (Sébastien and Bauler, 2013, p. 10).

Sirgy (2011) describes 6 theoretical concepts under which Quality of Life (QOL) indicators can be classified, among them, he includes "human development" and "socio-economic development". One of the main ideas stemming from his review is that understanding the theoretical basis of an indicators project will "help QOL researchers to develop and recommend relevant policies" (Sirgy, 2011, p. 18). Ideally, the theoretical framework should refer to theory and experts' inputs. The dimensions included in the index may also result

from societal debates[2], in which case disclosing the process may require a significantly extensive collection of documentation.

As was mentioned, authors of composite indices often face criticism due to the subjectivity -and possible arbitrariness- in their methodological decisions and assumptions. As justification, they should disclose evidence that the choices were guided in adherence to the theoretical framework. As Freudenberg (2003) affirms, the adopted theoretical framework will *ideally* allow the selection, combination, and weighting of variables in a way that accurately reflects the dimensions of the phenomenon intended to measure.

2.1.2 Data selection

The criteria for data selection, the relation with the conceptual framework, and sources should be clearly stated. If the utilization of proxy measures is needed due to the absence of data for specific units (countries), this approximation should be clearly explained. As suggested by the OECD (2008), a summary table on data characteristics, including availability, source, and type, should be shared by the authors.

In the case of variables dependent on "size-related" factors, such as GDP, the OECD (2008) proposes to scale the data to ensure comparability. If such transformation is needed, it should be stated and justified. Moreover, according to the phenomenon measured, there should be coherence regarding the consideration of input, output, or process indicators, or, if relevant, a combination of them (OECD, 2008). Additionally, the authors should describe if the considered indicators are objective, subjective, or a combination of these (UNECE, 2019).

2.1.3 Imputation of missing data

If imputations are needed, the authors should explain the methodology used to determine the values and justify the chosen method. Additionally, they should study the effect of the imputed data by examining the distributional characteristics of the series and the new values' coherency with the nature of the series (e.g., only positive values are meaningful) (OECD, 2008).

[2] UNECE (2019, p. 73) presents two examples of the use of societal debates to define the dimensions included in socio economic composite indicators.

2.2 Structural form of the index

2.2.1 Multivariate analysis

Authors should assess the degree of interrelation between the composite index' components by applying methods like Principal Component Analysis (PCA), cluster analysis, or the Cronbach's alpha coefficient (OECD, 2008). Moreover, whatever method chosen for multivariate analysis, authors should recognize the possible shortcomings of the tool.

Conducting multivariate analysis is an ideal step before choosing and executing weighting and aggregation methods appropriately (OECD, 2008), so it would be informative that such analysis is referred to in the index's methodological documentation. If possible, they should reveal the results (e.g. coefficients), which may also be the starting point for identifying areas where robustness checks were -or should be- performed.

2.2.2 Normalization

The chosen normalization technique should be stated and justified accordingly. Ideally, evidence should be provided about the implications that the chosen method may have in each dimension and on the index's overall results, as the method should be decided with respect to the data properties (UNECE, 2019). Authors may justify their choice by contrasting it with the alternatives and by running robustness checks. Some normalization methods involve using the data series own values to perform the scaling (e.g., z-scores standardization). However, if the normalization method involves the definition of minimum and maximum values as best- and worst-case scenarios, the process and rationale to determine these values should be disclosed accordingly.

Additionally, the chosen data series may have skewed distributions or include outliers. If the data exhibits these properties, they should be considered. Particularly, if transformations are performed before conducting the normalization procedure (OECD, 2008), such treatment should be disclosed.

2.2.3 Weighting and aggregation

Weighting*:* the selection of the weighting method should be justified, and the resulting weights should be disclosed. There are different approaches to weighting, sometimes classified in the range from objective to subjective. Objective approaches are characterized by using data-driven methods to determine weights (e.g., PCA or factor analysis). However, one could argue that such a process involves many assumptions, which are not deprived of

subjective judgment. On the other hand, subjective approaches rely mostly on experts' opinions in adherence to the theoretical framework.

Both approaches face criticism. Objective approaches are condemned for a supposed lack of conceptual meaningfulness. In contrast, subjective approaches are criticized for their proneness to the developer's bias and the room for inconsistencies, mainly due to the typical lack of a sound theoretical framework (Greco et al., 2019).

Nevertheless, there is no consensus on a specific method for achieving optimal weights. Essentially, *"*the absence of an "objective" way to determine weights and aggregation methods does not necessarily lead to rejection of the validity of composite indicators, as long as the entire process is transparent*"* (OECD, 2008, p. 33).

One of the key issues when discussing the usefulness and adequacy of using aggregate measures of achievements -or deprivations- in social sciences is the justification of the implied comparison between its components. A common practice is to assign equal weights across dimensions of the index, or within the components of each dimension. In that scenario, although the "method" to assign weights is entirely transparent, the offered justification should not go unnoticed. Such choice may be warranted as well as it just may be due to practical oversimplicity, bearing a cost on the actual value-added of the resulting index (Greco et al., 2019).

Another aspect related to defining weights that should be transparently accounted for is the underlying indicators' actual importance. As Becker et al. (2017) warn, pre-specified weights may be misunderstood as direct measures of the importance of dimensions and their underlying indicators. Assessing the degree of correlation between indicators would be essential, as double counting increases the dimension's actual weight (OECD, 2008). Ideally, authors should disclose the results of the correlation tests, and this can be achieved by estimating Pearson correlation coefficients, which Paruolo et al. (2013) denominate the "main effect" when socioeconomic variables are considered.

Therefore, care must be given on how the selected weights are communicated, as the indicators' actual importance is affected by the relation between the underlying indicators (correlation) and by other methodological choices (e.g., normalization and aggregation method). Studies such as those conducted by Becker et al. (2017) and Schlossarek et al. (2019) propose different methods to assess the components' importance.

Aggregation: regarding the aggregation method, authors should justify and explain its implications. Expressly, they should state if the aggregation method is compensatory or non-compensatory, shedding light on the allowed interplay of the index components performance. A discussion of different aggregation methods can be found in Greco et al. (2019).

The issue of allowing -or not- compensability between the index's components is critical when measuring multidimensional concepts such as human development. Depending on the underlying theoretical framework, an explanation of why underperforming in a dimension is allowed -or not- to be compensated by a better performance in another one should be provided. Arguably, the ample debate on the perfect compensability that was allowed between the dimensions of the HDI was one of the main reasons that prompted the change from arithmetic to geometric mean as a method of aggregation in 2010. Additionally, Munda and Nardo (2009) proved that when weights are defined as importance coefficients, authors should opt for a non-compensatory aggregation method for theoretical consistency.

Moreover, the chosen weights and the aggregation method will be fundamental as they influence the trade-offs between the index's components. The fact that composite indicators implicitly allow for the computation of trade-offs in measuring multidimensional socioeconomic phenomena, such as development, may give rise to questionable comparisons between the values of indicators that cannot be easily, if at all, compared. Decancq and Lugo (2009, p. 22) express that "the definite test for selecting a weighting scheme should be in terms of its reasonability in terms of the implied trade-offs between the dimensions". After analyzing different weighting methods, they conclude that reporting only the equal weights, without additional explanations, does not make explicit the underlying assumptions, as weights -along with data characteristics and the rest of methodological steps- define trade-offs between the index's components (Decancq and Lugo, 2009). Therefore, these trade-offs should be transparently discussed by the authors.

The Stiglitz-Sen-Fitoussi (2009) report stresses that, while creating composite indices, the process of weighting and aggregation put a relative value on the index components, affirming that the biggest problem is not explaining the weighting strategy but the often omitted explanation and justification of the normative implications of this process. A notable example can be found in Ravallion's (2012a, 2012b) research on the 2010 HDI, which yielded the unappealing conclusion of a much higher Marginal Rate of Substitution (MRS) of an extra year of life expectancy vs. income in rich countries than in those with low income.

2.3 Robustness tests

2.3.1 Uncertainty and sensitivity analysis

The conduction of robustness checks, mainly uncertainty and sensitivity analysis, is an important step in creating composite indicators to test the authors' assumptions, and it may lead to reducing the possibility of sending a misleading message (Saisana et al., 2005). Although these tests may be routinely and iteratively executed during the construction process, disclosing their results would contribute to explaining why the chosen methodology is considered superior compared to different alternatives. Moreover, reporting the index's results with associated uncertainty bounds would be a way of transparently accounting for the uncertainty related to the composite (OECD, 2008).

Therefore, disclosing this process entails explaining the alternative scenarios considered and the justification for their relevance. In the case of the selection of weights, Seth and McGillivray (2018) suggest a method for choosing alternative scenarios for robustness checks; their approach is based on the premise that there is a consensus about the maximum and minimum weight for each component of the index. Moreover, regularly disclosing robustness checks is a way of submitting the composite to a stricter standard. One possible way of reporting the sensitivity analysis is including scatter plots "with the values of the composite indicator for a country on the vertical axis and each input source of uncertainty on the horizontal axis" (OECD, 2008, p. 35).

2.4 Analysis of the index results against its components and related indicators

2.4.1 Back to the real data

Comparatively analyzing the resulting index against the original data series yields evidence of the aggregate measure's value-added. Transparently disclosing such comparison would be an argument in favor of using the composite instead of individually monitoring and analyzing its sub-indicators, which would be the equivalent of a dashboard approach.

The analysis may include correlation and causality tests. Additionally, explaining which component drives the country's performance would contribute to a correct interpretation of the results and would be essential for policy purposes. This analysis could be disclosed through a series of diagrams (e.g., leader-laggard decomposition, spider diagrams, traffic light representations, among others), accompanied by a brief description of the results (OECD, 2008).

2.4.2 Links to other indicators (or variables)

To enhance the explanation of the composite indicator and to test its meaningfulness, it would be useful to disclose the relation that the index has with other variables, principally through correlation analysis and, if possible, causality analysis (OECD, 2008). Therefore, authors should be careful in how they communicate the relation of the index results to specific contexts, transparently recognizing the sources of correlation without assuming causal explanations unless statistically proven.

Based on what has been described for each step, the different elements to be considered for the evaluation are summarized in Table 3.

Table 3 Summary of the framework for a transparency evaluation of composite indicators of development

Section	Step	Elements (to be disclosed)
Foundations of the composite indicator: theory and variables	Theoretical framework	- Definition of *what* is being measured and its objective - The description of its dimensions and, if applicable, other subcomponents
	Data selection	- Criteria for selection - Data characteristics and source - If necessary, detail of transformations (e.g., scaling)
	Imputation of missing data	- The methodology for imputation and its justification - Imputed values - Implications of the imputation on the index results
Structural form of the index	Multivariate analysis	- Description of the method used to conduct the analysis - Results of analyzing the relationship between underlying indicators
	Normalization	- Method used for normalization and justification - Implications of the method on the index results - Description of treatment for skewness or outliers - Additional information for the method's implementation (e.g., for a min-max approach: best and worst-case scenarios values and the rationale for setting them)
	Weighting and aggregation	- Weighting method used and justification - Weights assigned to each component of the index - Discussion of the relationship between weights and importance of the dimensions and their subcomponents - Aggregation method used and justification - Discussion about the allowed compensability -or not- between dimensions - Discussion about the implied trade-offs between dimensions, given the methodological choices
Robustness tests	Uncertainty and sensitivity analysis	- Sources of uncertainty and alternative scenarios considered - Results of analysis, including explanations and data visualizations
Analysis of the index results against its components and related indicators	Back to the real data	- Evidence of correlation and, if possible, causality analysis
	Links to other indicators (variables)	- Explanations of the theoretical relation between other indicators and the composite index - Results and discussion of correlation and, if possible, causality analysis

Source: author based on OECD (2008), UNECE (2019), and others cited in each corresponding section of this Chapter.

Chapter 3 Evaluation of transparency of composite indicators of development

In the present Chapter the selected composite indicators are evaluated according to the proposed framework. In each case, the assessment focuses on the latest available index methodology. As was mentioned, reference to previous versions is often necessary to provide a complete picture of the decisions made by the authors in the design and computation of each index, including justifications when provided. The structure of the evaluation follows the sections of the framework presented in Table 3.

3.1 Social Progress Index

About the index

The Social Progress Index (SPI) was developed by the Social Progress Imperative, an international non-profit organization. The index is calculated and published yearly from 2013[3] on the Social Progress Imperative website. The SPI is centered on social outcomes, as the authors affirm that their methodology is focused on isolating the non-economic dimensions of national performance (Stern et al., 2020). The summary of the SPI's structure is presented in Figure 1; a detail of its components can be found in Appendix A (Figure 1).

Figure 1 Summary of the Social Progress Index's structure

Source: author based on Stern et al. (2020).

3.1.1 Foundations of the composite indicator: theory and variables

Theoretical framework: the SPI was designed to measure social progress. The authors affirm that they constructed their definition in consultation with a group of academic and policy experts[4]: "social progress is the capacity of a society to meet the basic human needs of its citizens, establish the building blocks that allow citizens and communities to enhance

[3] The index was created in 2011 and a report with a beta version for 50 countries was published in 2013. On their website (socialprogress.org), they publish the results and methodology reports from 2014.
[4] In the methodological reports from 2014-2017 a list of experts is acknowledged as part of the consultancy team that led to establishing the theoretical framework. Additionally, an extensive list of sources is provided in the bibliography of each year's report; however, not all the sources are cited on the main body of the report.

and sustain the quality of their lives, and create the conditions for all individuals to reach their full potential" (Stern et al., 2020, p.4).

This definition is grounded in Amartya Sen's capabilities approach, cited in the methodological reports from 2014 to 2017. Although from 2018 onwards Amartya Sen's *Development as Freedom* (1999) and *Commodities and Capabilities* (1985), and the Stiglitz-Sen-Fittousi (2009) report are not quoted directly to support the social progress' definition, they are included as proposed bibliography at the end of every year's methodological report, among other relevant sources.

The dimensions included in the index are three: Basic Human Needs, Foundations of Wellbeing, and Opportunity. These dimensions are intuitively directly linked to the propositions contained in the social progress definition. In the 2014 report, a chapter is dedicated to discussing the conceptual foundations of the SPI; however, it is primarily centered on the philosophical grounding of the index and recognizing the normative essence involved in its construction (Stern et al., 2014).

The dimensions and their components are described as derived from Natural Rights philosophy and referenced with sources from classical political economy, ethics, and human rights theory (Stern et al., 2014, pp. 36-38). However, the SPI's dimensions and components descriptions are not immediately accessible: dimensions are only briefly described in the 2020 methodological report, and for components only one line of text per component is included (Stern et al., 2020, p. 4). A slightly longer description by component can be traced to the 2017 SPI's main report (Porter et al., 2017, p. 16-17).

As suggested by Stern et al. (2020), the index is intended to be used by the public to "compare their country against others on different facets of social progress, allowing the identification of specific areas of strength or weaknesses" (Stern et al. 2020, p. 3). For this purpose, they publish the index results through country specific scorecards, which contain the scores attained and the relative position (ranking) for each indicator, component, dimension, and the overall SPI score (an example of a scorecard is presented on Appendix B).

Data selection: the choice of indicators is guided by an indicator selection tree which can be consulted in Appendix C of this document. One must point out that the selection criteria presented in the referred tree are not always clear. Notably, for the guiding question "Does this indicator measure an input or an outcome?", the options are either considering it an outcome indicator because it is: "A concept that we are interested in because it is good or

bad for its own sake"; or "Important mainly because it signals something else and is therefore an input indicator" (Stern et al., 2020, p. 8); arguably, this is a confusing way to distinguish between two categories of indicators.

The authors remark that some high-quality indicators have not been used due to the low number of countries for which the data is available. Year to year, they publish the changes in the indicators used in the methodological report. The complete list of indicators, with their characteristics (definition, unit, and source), is included in the 2020 report by Stern et al. (2020, p. 20).

The authors affirm that the SPI is an outcome-based measure; this was decided following the index's objective: capturing the performance of non-economic indicators. However, it may be argued in some cases that the indicators could also be considered inputs rather than outcomes (i.e., primary enrolment rates).

The raw data used to calculate the 2020 SPI is available for download on the index website (Social Progress Imperative, 2020). Several indicators suffer transformations[5], which are performed to enable comparability between countries and, in some cases, to avoid the influence of outliers. The methodological report enlists all transformed indicators and the reasons that motivate them (Stern et al., 2020, p. 9).

Imputation of missing data: the values imputed before starting the index calculation are included in the raw dataset, along with the method employed to obtain them[6]; however, as claimed in the methodological report, other values are imputed through regression analysis during the calculation of the index, and the estimated values are not disclosed.

Stern et al. (2020) affirm that whenever there is a missing value for an indicator, they assess the imputation methodology "both before and during index calculation"[7] (Stern et al., 2020, p. 11). They provide examples of how they assess the accuracy of the predicted values, for example, by imputing values for countries missing just one indicator per component. Based on the previous statements, it could be argued that the authors have considered the effect that imputed values have throughout the calculation process.

[5] These include capping, log transformation, and "calculation of parity". The latter refers to a specific case in which they propose the value of 1 for gender parity in school enrollment, and then proceed to estimate the distance to it per country.
[6] The legends "Interpolated", "Future value used", and "Historical value used" are inserted as comments in each of the data cells corresponding to a specific country and year.
[7] When imputations are done through linear interpolation, they affirm that the method is favored to ensure that improvements -or declines- in performance are not due to the gaps in the data (Stern et al., 2020, p.11).

3.1.2 Structural form of the index

Multivariate analysis: given the statistical properties at the core of the SPI's methodology, it is possible to confirm through various anecdotal propositions that the authors recurrently performed multivariate analysis to determine the index's structure. For example, while explaining the reason why many of the weights within components are similar, they mention that this is because they "ensure a fair level of correlation between them (e.g., not too high or low correlation) prior to finalizing our framework" (Stern et al., 2020, p. 13).

Additionally, they dedicate a section of the methodological report entitled "Structural Integrity of the Social Progress Index" to different tests that intend to validate the factor analysis. Though the correlations between the indicators are not disclosed, they establish the range for which they consider the indicators to have this "fair level of correlation" (between r=0.3 and r=0.92) (Stern et al., 2020, pp. 15-16). The conceptual fit between indicators within each component is evaluated using Cronbach's alpha coefficients, which must be above 0.7 to be valid. These coefficients are disclosed in the report, and the inclusion of the indicators within the component Environmental Quality, which do not fulfill the above 0.7 requirement, are justified on the grounds of a lack of alternatives to represent the component. Lastly, they test the goodness of fit within components by estimating the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy.

The disclosure of these tests is informative and leaves evidence of the efforts to exhibit statistical coherence. However, since the complete dataset, including all the values of imputations performed during calculation is not provided, it is not possible to replicate the exercise with the available information.

Normalization, weighting, and aggregation: normalization takes place at two levels: indicators and components. At the indicator level, the method used is z-score standardization, and, for components, a min-max approach is used.

First, they set best and worst-case scenarios for each indicator "to provide concrete boundaries" (Stern et al., 2020, p. 12). The setting of these scenarios is guided by theoretical and historical values (Stern et al., 2020). Sometimes the boundaries are given by the natural scale of the indicator (e.g., an indicator is the Global Peace Index, which is already measured in the range 0-5). In other instances, the boundaries are set using historical information of the worst (or best) values recorded five years before 2011, the first year for which they estimate the SPI with the 2020 methodological framework. These extreme values are

disclosed in the report (Stern et al., 2020, p. 32), nevertheless, the rationale for setting such scenarios is not disclosed on an indicator by indicator basis. Additionally, they invert 21 indicators that measure adverse outcomes, so a higher value denotes a better situation. All inverted indicators are enlisted in the methodological report (Stern et al., 2020, p. 12). After setting these boundaries, they perform a z-score standardization.

Applying z-scores as a normalization strategy at the indicator level is necessary to conduct Principal Component Analysis (PCA). Essentially, PCA assigns weights to the indicators within a component, "capturing the maximum amount of variance in the data while reducing redundancy between indicators" (Stern et al., 2020, p. 13). The justification for favoring PCA-generated weights over equal weighting at the indicator level is that they wanted to ensure that "indicators are meaningfully contributing to a component score while accounting for similarities between them" (Stern et al., 2020, p. 13). The authors reveal the software and command they use to generate the weights through the statistical software (Stern et al., 2020, p. 13). The weights generated for each of the indicators are disclosed in the report by Stern et al. (2020, p. 34-35).

After, they generate component values as a weighted average by multiplying each weight with its corresponding indicator. Subsequently, to generate the component scores, a min-max approach is performed by considering the maximum and lowest values achieved among the countries for which the SPI is calculated. Although these bounds are also referred to as best and worst-case scenarios (Stern et al., 2020, p. 13), their setting is guided by the actual highest and lowest component values estimated for the sample.

The aggregation of every set of four components, and of the three dimensions to the overall index, is performed through arithmetic average, which means that they set nominal equal weights both at the component and dimension level[8]. In the methodological reports from 2014 to 2017, they claim that the equal weighting at the component and dimension level is guided by the absence of logic for different weights.

The justifications offered by the authors can be traced to the following statements: "(…) the Social Progress Index architecture equally weights components for constructing a dimension-level score because there is no clear theoretical or empirical reason to weight any of the

[8] On their website, in the section "Global Index – Methodology" they claim that the explanation for choosing equal weighting, and other alternatives, are discussed in the 2018 methodological report; however, such account has not been found. Conversely, some alternative weights scenarios are discussed in the 2017 methodological report of the index (Stern et al., 2017, p. 32).

components more highly than any other" (Stern et al., 2014, p. 8; Stern et al., 2015, p. 8; Stern et al., 2016, p. 9; Stern et al., 2017 p. 7).

Though anecdotal, there is some evidence of alternative considerations for equally weighting components and dimensions: "We considered other avenues to weighting such as using the coefficients of a regression of life satisfaction scores against the three-dimension scores. Though the results are intriguing (and an avenue we intend to explore in ongoing work), we did not believe there was a sufficiently robust relationship of how each of the social progress dimensions mattered in a relative way. We therefore adopt a simple average of the dimensions in order to highlight the critical role of each in social progress" (Stern et al., 2014, p. 7; Stern et al., 2015, p. 7; Stern et al., 2016, p. 8; Stern et al., 2017 p. 7).

Overall, no comments are made regarding the compensability allowed between the index components and dimensions by favoring arithmetic mean as aggregation method. Therefore, there is an absence of discussion of the implied trade-offs between the index components and dimensions.

3.1.3 Robustness tests

Uncertainty and sensitivity analysis: in the 2017 report, the authors considered alternative methods to some of their methodological choices. First, they applied PCA at the component level and concluded that the index is multidimensional in the sense that the components jointly contribute to a common factor, both within dimensions and to the overall index. Applying the same analysis to each dimension, they concluded that "the contributions of each component are relatively balanced" (Stern et al., 2017, p. 32).

Secondly, they tested the equal weighting at the component level by removing one component at a time and assessing the changes in the country ranking, concluding that the changes were moderate, except for the case of two components that were believed to have "slightly more influence" (Stern et al., 2017, p. 33) on the overall ranking. Nevertheless, they did not consider a full set of alternative weights at this level. Lastly, they tested the effect of applying equal weights at the indicator level, which prompted drastic changes in the ranking of the Opportunity dimension. Therefore, they conclude that PCA "can help correct for the relationships between indicators, while preserving the framework's conceptual basis" (Stern et al., 2017, p. 33).

Additional to the author's analysis, the European Commission's Joint Research Centre (JRC) conducted a sensitivity analysis for the 2018 SPI, which is mentioned in the 2018

methodological report (Stern et al., 2018, p. 21). They ran 2000 different simulations to test how the ranking of the 2018 SPI varied on two of the methodological choices made by the authors: the equal weighting at the dimension level and the arithmetic aggregation at the dimension level that leads to the overall index[9]. The conclusion was that the equal weighting at the dimension level and the arithmetic aggregation were robust to different choices. (Norlén and Caperna, 2018). Nevertheless, the authors do not disclose sensitivity analysis for the 2020 methodology of the SPI, whose indicator framework differs from the one previously analyzed by the referred studies.

3.1.4 Analysis of the index results against its components and related indicators

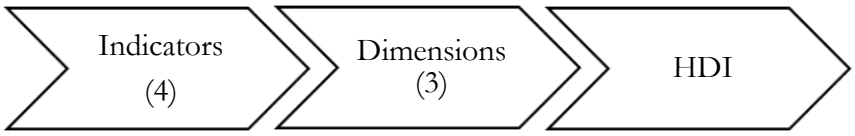
Back to the real data and links to other indicators (or variables): as mentioned before, the authors test the fit of the indicators to explain the components through the Cronbach's alpha coefficient and the KMO. The JRC, during the statistical audit, performed and disclose a correlation analysis of the SPI and its components, in which they conclude that high correlations denote a possible redundancy at the component level of the index, thus suggesting lowering the number of components (Norlén and Caperna, 2018). However, for the 2020 version of the index, they do not disclose an analysis contrasting the SPI results with its underlying components or other composite indicators or variables.

3.2 Human Development Index

About the index

The HDI was designed by Mahbub ul Haq and published for the first time in 1990 by the UNDP. The measure acquired much popularity and is arguably the most recognized composite index in the development field. The summarized structure of the index is portrayed in Figure 2, and a detailed representation can be found in Appendix A (Figure 2).

Figure 2 Summary of the Human Development Index's structure



Source: author based on UNDP (2020b).

[9] They tested 1000 different weighting schemes, all generated through Monte Carlo simulations. The different weights were in turn combined with aggregation done through either arithmetic or geometric mean, totaling 2000 simulations. Alternatives to the indicator PCA weights were not simulated as they would introduce "uncontrollable changes" (Norlén and Caperna, 2018, p. 19) to the index's structure; although they did point out that PCA weights were quite similar to equal weights.

The HDI is published each year as part of the Human Development Report (HDR). The HDR also presents other indicators and composites that provide a more comprehensive picture of global development. The HDI methodology was routinely published as a synthetic Technical Note embedded in the main document of the HDR until 2011. Since 2013, the yearly Technical Note is published as a separate document and directly on the UNDP-HDR website[10]. In the 2011 and 2013 Technical Notes, there is an indication referring the reader to the paper by Klugman et al. (2011) for "a full elaboration of the method and its rationale" (UNDP, 2011, p. 168; UNDP, 2013, p. 2); this indication is not reproduced in the Technical Notes from 2015 to 2020[11].

Many papers by UNDP-related authors -and external critics alike- have been published discussing the HDI's properties and shortcomings. This high exposure and scrutiny provide the UNDP with opportunities to evaluate alternatives, which have prompted several methodological modifications throughout the years. Although many suggestions are deemed inappropriate or unfeasible, various critiques are mentioned in the 2020 HDR, which offers different points of view and sheds light on the index's limitations. A summary of various of the critiques faced by the HDI can be found on Kovacevic (2011).

3.2.1 Foundations of the composite indicator: theory and variables

Theoretical framework: the UNDP grabs from Amartya Sen's capabilities approach to define human development. In its first edition, human development was described as "a process of enlarging people's choices" (UNDP, 1990, p. 10); after 30 years, the definition has slightly changed to highlight the subjectivity that is at the core of the concept "enlarging people's abilities and opportunities to be and do what they have reason to value" (UNDP, 2020a, p. 21).

Novel as this proposal was, the translation of such a broad concept as *capabilities* to just three dimensions with one indicator under each of them — Long and Healthy Life, Knowledge, and A Decent Standard of Living — came with the clarification that "the measurement of human development should for the time being focus on the three essential elements of human life" (UNDP, 1990, pp. 11-12). However, to this date, the same three dimensions remain, and the HDI is proposed as a "summary measure of achievements in three key

[10] hdr.undp.org.

[11] Nevertheless, the HDI maintains to a great extent the structure introduced in 2010, which is the one that Klugman et al. (2011) discuss, reason for which it will be cited in this study when relevant.

dimensions of human development" (UNDP, 2020b, p. 2)[12]. As Lustig (2011) comments, there are measuring difficulties and political controversies regarding the possibility of including other dimensions, such as political freedom or agency.

Data selection and imputation of missing data: the choice of indicators for the education and health dimensions was guided by the belief that they indirectly hold information about other relevant indicators within each specific dimension. In the case of the education dimension, various critiques have pointed out that the proposed indicators lack the assessment of educational quality – which can also be suggested of the longevity indicator. On the other hand, the inclusion of the Gross National Income (GNI) per capita as the standard of living indicator was guided by the notion that income is a necessary means to fulfill capabilities other than health and education (UNDP, 2020b). The variety of critiques regarding indicator choices have elicited further justifications for their inclusion; some of them are summarized in García and Kovacevic (2011)[13].

The raw data for the calculation of the index is regularly published on the HDR's Statistical annex. The dataset of the 2020 HDI is also available for download in .xlsx format (UNDP, 2020c). For missing values, various notes are included on the .xlsx file to point out to the source of the data or the imputation strategy used to obtain the values presented. Regarding transformations, the income variable, GNI per capita, is used in its logarithmic form to reflect that, as income grows, the degree to which it can be converted into capabilities decreases (UNDP, 2020b). As mentioned before, this transformation contributed to an unfortunate consequence summed up by Ravallion (2012a, 2012b), who demonstrated that a troublesome relative comparison was generated in the 2010 HDI's methodology: that extra years of life in low-income countries are less valuable than in rich countries.

3.2.2 Structural form of the index

Multivariate analysis: through various documents published by the UNDP, it can be confirmed that the relation between the three dimensions of the HDI has been studied at several points. Kovacevic (2011, p. 17) presents a summary of correlations between the index components and the overall HDI score for the 2009 HDI, determining that, although

[12] Moreover, since 1990 the UNDP has expressly said that, even though there are evidently more variables to consider with respect to human development, "too many indicators could produce a perplexing picture" (UNDP, 1990, p. 11). Thirty years after its creation, though several changes both in the indicators and on the structure of the index have been made, the number of indicators increased just by one.

[13] This paper is cited by Klugman, et al (2011), specifically to justify the robustness of the HDI to different weights.

correlations are high, the Cronbach's alpha coefficient signals the internal consistency and multidimensionality of the index. However, the education indicators changed in 2010. For the current set of indicators, a complete analysis and disclosure of correlations have not been found on the HDRs, the related Technical Notes or other publications by the Human Development Report Office (HDRO).

Nevertheless, García and Kovacevic (2011) point out to the high correlation between dimensions as a reason for the robustness in the choice of equal weights, and cite Kovacevic (2011) as a proof of non-redundancy, even when those calculations correspond to an older indicator framework. Additionally, from the HDRO perspective, the possible redundancy has been dismissed as inaccurate through a correlation analysis of the changes -rather than the level- of both HDI scores and its underlying indicators (Klugman et al., 2011).

Normalization: the normalization procedure is done through a min-max approach with fixed goalposts. To perform this method, the UNDP establishes best and worst-case scenarios. From 1990 to 1993, these scenarios were the actual highest and lowest values found among the respective dataset; however, to enable comparisons over time, the goalposts were fixed since 1994 (UNDP, 1994). Variable upper bounds were adopted in 2010 but were abandoned since they affected intertemporal comparison (Klasen, 2018). The current goalposts have been in place since 2014.

The rationale for setting the bounds at their current levels is briefly explained in the 2020 Technical Note. A theoretical argument is used to set the maximum value for GNI of 75,000: citing Kahneman and Deaton (2010), it is considered that, above that income, "there is virtually no gain in human development and wellbeing" (UNDP, 2020b, p. 2)[14]. For the rest of the bounds, these correspond to either historically observed or estimated values, each accompanied by an explanation (i.e., lower bounds in education are set at 0 since it is deemed that societies can conduct their lives without formal education).

As García and Kovacevic (2011) mention, another possible normalization technique would be z-scores; however, they claim that this strategy would reward exceptional performance, which would enable some degree of compensability among the index dimensions.

[14] It is worth noting that Kahneman and Deaton (2010) conducted the research regarding subjective well-being in the United States (US). Since the income indicator is supposed to cover a wide variety of other objective capabilities, more explanation should be warranted with respect to it. Moreover, a recently conducted research on well-being of employed population in US found no evidence of a plateaued experienced wellbeing above $75,000 (Killingsworth, 2021).

Additionally, they comment on how min-max allows for differentiation between countries with similar levels of achievements due to how these methods widen the index range.

Weighting and aggregation: HDI's dimensions are weighted equally based on the idea that each of them has the same worth. However, the team of the HDRO has recognized that the actual importance -and contribution- of a dimension to the index is linked at least to the distribution of the variables and the normalization method (Kovacevic, 2011; García and Kovacevic, 2011).

Addressing critiques to the full compensability allowed between dimensions of the HDI from 1990 to 2009, the geometric mean has been employed as an aggregation method since 2010. When first released, the UNDP highlighted the superiority of the new aggregation formula through various statements, for example, by suggesting that it allowed "poor performance to be directly reflected in the HDI" (UNDP, 2010, p. 15) and that "as a basis for comparison of achievement, this method is more respectful of the intrinsic differences in the dimensions than the simple average is" (UNDP, 2010, p. 15). Anand (2018), among other critiques, raises the issue of how unclear the previous extracts are and mentions how the old version of the HDI (which employed arithmetic mean) was also able to reflect the poor performance of the underlying indicators.

As mentioned, the effect of the geometric aggregation and the log transformation of income have been the focus of critiques by Ravallion (2012a, 2012b), and this critique has been acknowledged by the HDRO. Klugman et al. (2011) commented how Ravallion's interpretation of the trade-offs, calculated as Marginal Rates of Substitution (MRS), is based on the wrong assumption that the HDI is a welfare function that should be maximized, which would imply that the MRS can be understood as relative values (how much is an extra year of life worth).

They dismiss the critique by stating that the HDI is instead a capabilities index, which does not intend to contain a complete account of capabilities, and for this reason, should not be subject to maximization, as not every relevant aspect for development would be addressed by improving HDI scores. Therefore, they affirm that the MRS commented by Ravallion can "be better interpreted as reflecting different contributions of income vis-à-vis health in furthering capabilities" (Klugman et al., 2011, p. 278). Ravallion (2011) counterargues that the debate should not be centered in determining if the HDI is a complete metric, or if it should be perceived as a development objective in itself, but that the MRS of composite indicators is in itself important given that "it is the relative weights on its dimensions that

matter the most" (Ravallion, 2011, p. 477). Although the critique has been dismissed by the UNDP, the persistence of the debate is expressly recognized in the 2020 HDR, along with a brief discussion of alternatives proposed by other authors (UNDP, 2020a, p. 247).

3.2.3 Robustness tests

Uncertainty and sensitivity analysis: different methodological choices concerning the HDI have been tested and disclosed through background papers published by the HDRO, although the analysis corresponds to the 2010 HDI. Gidwitz et al. (2010) performed a sensitivity analysis for different imputation methods, and their results are cited in the 2010 HDR (UNDP, 2010, p. 207). They concluded that the methodology employed for this purpose by the HDRO is robust, as the other techniques did not yield significantly different results than those employed by the UNDP.

García and Kovacevic (2011) performed a sensitivity analysis for several scenarios. The choices subjected to this analysis were: weighting, the functional form of life expectancy (log transformation), and the minimum goalposts chosen for the normalization strategy. They disclose the statistical reasoning and results for every choice. Ultimately, they conclude that the HDI is a robust measure and shed light on the data's inherent limitations. Though this is a highly technical and comprehensive study of the HDI, this type of analysis is not regularly performed by the UNDP and has not been found for the 2020 HDI.

3.2.4 Analysis of the index results against its components and related indicators

Back to the real data: correlation analysis of the HDI and its underlying indicators has been conducted and presented for previous years. García and Kovacevic (2011) tested the relation between variation in the HDI's dimensions and the HDI results by estimating Pearson correlation coefficients. Kovacevic (2011) found that the income dimension is the one that mainly explains the variation in the HDI; this is one of the properties that makes it evident that, even when equal nominal weights are set, these are not directly representative of the importance that each dimension has in the index. A more recent version of this type of analysis performed by the HDRO/UNDP has not been found.

Links to other indicators (or variables): throughout the years, the relation of the HDI with other indicators, such as GDP, has been studied by the UNDP and external authors. Klugman et al. (2011) analyze the change in HDI and income growth, finding evidence of strong correlations; this is because income is the dimension that mainly drives the HDI

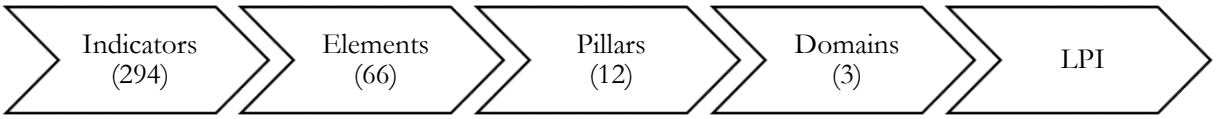
results. However, a section of the 2010 HDR is dedicated to discussing the finding that there is a low correlation between improvements in per capita income and improvements on an HDI calculated just with its non-income dimensions (UNDP, 2010; Klugman et al., 2011).

3.3 Legatum Prosperity Index

About the index

The Legatum Prosperity Index (LPI) was created in 2007 by the Legatum Institute and is published yearly. The LPI's structure, summarized in Figure 3, comprises a series of levels: it counts with three domains, each containing four pillars. Groups of policy-focused elements represent the pillars. To capture the elements, 294 indicators from a variety of sources are used. A detailed diagram of the index's components can be found in Appendix A (Figure 3).

Figure 3 Summary of the Legatum Prosperity Index's structure



Source: author, based on Legatum Institute (2020a).

3.3.1 Foundations of the composite indicator: theory and variables

Theoretical framework: the LPI was designed to measure prosperity, defined as "when all people have the opportunity and freedom to thrive" (Legatum Institute, 2020b, p. 2). Moreover, they affirm that their research and consultations with 100 experts in different fields[15] make it apparent that prosperity is multidimensional.

The index's three domains are Inclusive Societies, Open Economies, and Empowered People. The Legatum Institute (2020b) dedicates a report to explain what they try to capture through each domain, pillar, and per element. To support the descriptions, references to either technical studies or theory are provided at the domain level.

Data selection: the 2020 LPI comprises 294 underlying indicators, whose selection is guided by a set of criteria and advice from experts. This selection process involves assessing the relationship with the element they represent, which is done through "conceptual and statistical reasoning" (Legatum Institute, 2020c, p. 5). They affirm that each indicator is

[15] They provide a full list of advisors classified by pillar for the formulation of the 2020 LPI, including a brief profile in the index webpage (Legatum Institute, 2020d).

supported by the academic literature "with a wide consensus that they capture the underlying meaning of the element" (Legatum Institute, 2020c, p. 5); even when they describe each component, one cannot find references to literature addressing each element and to such *wide consensus*. Nevertheless, on the data description tables, they provide clarifications for some indicators, sometimes justifying their inclusion (Legatum Institute, 2020c, pp. 20-60).

Cronbach's alpha coefficients are evaluated as a measure of fit of indicators within an element. Additionally, as a criterion for including indicators that are "plausibly a causal factor of both wealth and wellbeing" (Legatum Institute, 2020c, p. 5) they test correlations of each indicator with two proxy measures: productive capacity and Cantril's ladder. Productive capacity is a concept they estimate as GDP minus resource rents and divided by working-age population[16]. Cantril's ladder, corresponds to an ordinal variable, coming from the 2018 Gallup's World Poll for 140 countries, which measures self-reported life satisfaction by asking people to report on which step (0 worst life -10 best life) they currently think they are on.

Other criteria considered are spatial and temporal coverage to maximize the number of countries for calculating the LPI. Though they provide an account of changes in the included indicators from year to year, and a few notes describing indicators that would be more suitable for measuring some elements but had to be discarded for different reasons[17], they do not provide a complete list explaining which indicators were discarded.

The methodological report includes a list with the essential characteristics of the data (description, unit, source, and last update), and the entire dataset for the calculation of the 2020 LPI is downloadable in .xlsx format (Legatum Institute, 2020e). The database includes both the raw data and the score per indicator, element, component, and the overall index from 2007. These sets of scores provide an overview of the relative position of each country in a large set of variables[18]. In 44 cases, which are fully reported, they perform log transformations to minimize the effect of outliers in the data (Legatum Institute, 2020c).

In the section where they briefly discuss LPI's limitations, the Institute affirms the following: "There are always challenges obtaining data that captures the core idea of what we are trying

[16] A small essay explaining why they considered this a proxy of wealth is included on the 2019 LPI's report (Legatum Institute, 2019a. pp. 52-55)

[17] For example, one of the notes included is "Our expert working group indicated the potential importance of including data on gang related crime within this element, but the lack of suitable data prevented us from doing so" (Legatum Institute, 2020c, p.22).

[18] They also disclose these scores on Country Profiles that can be downloaded directly from their webpage (Legatum Institute, 2020f).

to communicate. That is why, in some cases, we need to use outcome data rather than input data" (Legatum Institute, 2020c, p. 15). Qualifying the use of outcome indicators as a limitation is confusing given the LPI's aim and framework, in which they specifically state that "By using the Index, it is possible to compare the relative performance of each country for overall prosperity" (Legatum Institute, 2020a, p. 4). Though this is offered as a general limitation of the index, it may be possible that they refer only to a few cases; however, they do not indicate in which instances this type of compromise had to be made.

Imputation of missing data: to complete the dataset, they employ one of three distinct methods. Firstly, if possible, they use the latest -or earliest- known value of the variable. Secondly, the Institute performs "augmentations" of data: they search the same variable in another source, which may give rise to issues of comparability that are not addressed in the report. Additionally, if a variable is only available for some countries, but they deem that it may be defined for others "by virtue of the source's methodology" (Legatum Institute, 2020c, p. 6) they allocate an assumed value[19]. Lastly, they conduct imputations through linear regressions involving the utilization of different variables to approximate the missing values (e.g., productive capacity, country groupings, or other driver variables).

It is important to note that a complete account of the final imputed values, the technique used to generate them -and the process itself- are not provided. Instead, they only reveal the percentage of values imputed per pillar for those countries with 15% or more indicators imputed (Legatum Institute, 2020c, pp. 67-68). From this incomplete account of imputations, one can observe that the degree of imputation by country and pillar is higher than 50% in many cases (e.g., the imputed values for 27 out of 40 countries in the Governance pillar is over 50%).

The LPI is not calculated for countries with an overall 50% or more indicators missing; however, an imputation level of 25% would mean that the authors generated 70 out of 294 indicators. Therefore, the scores of the LPI should be carefully examined for countries with high levels of imputed values. Moreover, there is no discussion about the effect of the high amount of imputations on the index.

[19] They provide one example for an indicator (Bertelsmann Transformation Index) that was only available for developing countries, so for the OECD countries that are not included the Institute imputes the best possible value (Legatum Institute, 2020c, p. 6). This practice may be considered obscure, and each case should be clearly reported.

3.3.2 Structural form of the index

Multivariate analysis: multivariate analysis is performed at two levels of the index: as cited, they evaluate the Cronbach's alpha coefficients as one of the selection criteria for indicators within elements; and coefficients are also calculated for elements within a pillar as a measure of internal consistency. Nevertheless, only the coefficients at the pillar level are revealed in the methodological report (Legatum Institute, 2020c, p. 65).

Although the Institute confirms that they adopt a rule of thumb of a Cronbach's alpha coefficient to be of 0.7 or above to ensure internal consistency, they confirm that there are two pillars (Social Capital and Natural Environment) and an undisclosed number of elements that do not fulfill this requirement. Nevertheless, they opted to maintain them since, after discussing it with external experts, they determined that "reasons for their inclusion counterbalanced the statistical findings" (Legatum Institute, 2020c, p. 13). None account of this discussion is found in the LPI's documentation, and not even examples are provided of elements that were not internally consistent.

Normalization: the normalization method used is a Distance to Frontier approach (DTF)[20] in which they compare each country's performance with assumed best and worst-case scenarios for each indicator. These scenarios are defined considering either historical data or "logical frontiers" when the indicator already moves within a pre-specified range (e.g., the indicator "budget transparency" is an index that ranges from 0-100) (Legatum Institute, 2020c).

For most cases, the authors affirm that they try to set the frontiers so that the normalized values contain a relatively similar standard deviation across the indicators (Legatum Institute, 2020c). A list of these best and worst scenarios per indicator is not provided; thus, the rationale that led to their setting is not apparent for all indicators.

From 2007 until 2015, the chosen normalization method was z-score standardization. Though the change to a DTF method was a significant methodological change, it has not been possible to trace a discussion about it on the available reports. Nevertheless, on the first report where the DTF was employed, the Institute affirms that the DTF enables overtime

[20] Though referred to with a distinct name, this method uses the same formula as a min-max approach with fixed goalposts.

comparisons of the results, which may be one of the reasons guiding the change (Legatum Institute, 2016).

Weighting and aggregation: the chosen method for assigning weights to indicators is based on the Institute's opinion on their relevance to describe the element to which they pertain, based on "the academic literature, policy debate, and expert opinion" (Legatum Institute, 2020c, p. 8), reliability, and its significance with economic and social well-being in a global context (correlation with productive capacity and Cantril's ladder). The weights are assigned as factors of either 0.5, 1.0, 1.5 or 2.0[21].

Based on the same criteria, they weigh the elements by assigning percentages that they claim reflect the relative importance of the pillar and its contribution to prosperity. Therefore, weights are explicitly labelled as measures of the "importance"[22] of the indicator to the element (and of the element to the pillar) they belong to in each case. Moreover, on their website they explicitly claim that "Each indicator is assigned a weight, indicating the level of importance within the element it has in affecting prosperity" (Legatum Institute, 2020g). Qualifying these nominal weights as importance is misleading, as each indicator's actual effect on prosperity (LPI) is at least affected by the statistical properties of the data and the choices of normalization and aggregation.

Pillars are equally weighted, as well as the domains to which they belong. In aggregating from pillars to domains and from domains to the LPI score, the arithmetic mean is employed (Legatum Institute, 2020c). No justification is provided for the choice of aggregation method. Although the authors are adamant in describing why it is essential to reflect the difference of importance between indicators and elements through distinct weights, they do not discuss the compensability allowed by the arithmetic aggregation at the pillar and domain levels. No discussion is offered about the implied trade-offs between the components of the index.

[21] No explanation is offered for how this scale of weights was formulated, although one can presume that the 0.5 intervals reduce the difficulty of simultaneously assess the relative importance of various indicators (i.e. an indicator is twice as important as other).

[22] "each indicator is assigned a weight within an element, indicating the level of importance it has in that element. Similarly, each element has a weight that reflects its importance in the overall pillar" (Legatum Institute, 2020c, p.8).

3.3.3 Robustness tests

Since 2016, the Institute routinely performs sensitivity analysis to test the choice of weights (Legatum Institute, 2016, p. 20; 2017, p. 20; 2018, p. 20; 2019b, p. 32; 2020c, p. 13). Other choices, such as normalization and aggregation, are left out of this analysis. The alternative scenarios tested for weighting are a) equal weights at all levels of aggregation and b) 1000 sets of randomly allocated weights of 0.5, 1.0, 1.5, or 2.0 through Monte Carlo simulations. In both types of scenarios, they conclude that the relevant differences in ranking were mainly due to indicators that they had purposely down-weighted due to lack of data (Legatum, Institute, 2020c, p. 13).

Therefore, the Institute judges that the LPI's weights, which are ultimately defined by experts' opinion, are meaningful to account for the lack of data (Legatum Institute, 2020c). This affirmation raises an issue: even when data reliability is one of the criteria for defining weights, it may be that the down-weighting of indicators is misinterpreted as the minor relative importance of an indicator within the element[23]; therefore, it would be particularly important to provide the rationale for down-weighting in such instances.

3.3.4 Analysis of the index results against its components and related indicators

Back to the real data: with regards to the LPI's underlying data, a systematic review of what indicators, elements, or pillars principally drive the LPI results is not provided. Although the Pearson correlation coefficients between each pillar and the overall index are disclosed in the methodological report Appendix, the authors do not discuss these results (Legatum Institute, 2020c, p. 61-64). Only for a few countries, those identified as outliers while performing the sensitivity analysis, are some pillars identified as drivers of prosperity.

Links to other indicators (or variables): to test the index's relevance in relation to other concepts, the Institute analyses how the LPI may influence the two variables that according to them capture wealth and wellbeing: productive capacity and Cantril's ladder. However, while selecting the underlying indicators, the Institute ensures that each of them is statistically correlated with these same two variables; therefore, it should be commented that these correlations were already expected from earlier analysis.

[23] An example is that of pre-primary enrolment, which is down-graded due to lack of data "despite it perhaps having as much importance as other levels of education" (Legatum Institute, 2020c, p. 13). Unless a full account is provided it is not clear for which indicators this rationale has been applied.

Since 2019, the results of the LPI are also contrasted with those of the SPI, HDI, and the Global Competitiveness Index (GCI); the Institute considers these composites as they provide measures of specific components of prosperity. The results are portrayed as simple regressions of the LPI results vs. each index, through which they conclude that correlations among them are high. They also provide brief explanations for a limited number of outliers identified in the regression (Legatum Institute, 2019b p. 34; 2020c, p. 15)

Chapter 4 Comparative discussion of the evaluated composite indicators

The present chapter offers a comparative discussion of the transparency with which the three composite indicators are disclosed. The analysis is sustained on the evidence disclosed on the previous Chapter. As mentioned, the discussion's objective is to highlight identified omissions -or just partial recognition- of important considerations about the indexes' construction process and properties.

As an initial consideration, it is undeniable that, while disclosing the methodology for the HDI, the UNDP benefits from the inherent simplicity of its structure. Additionally, given its longevity and its high-profile presence, the HDI has been subject to high scrutiny, thus eliciting the response of the UNDP in acknowledging and justifying their decisions and their implications, though these recognitions may not be updated and immediately accessible[24]. Nevertheless, the SPI and LPI were designed by renowned scholars, and benefit from the advice of experts in each of the dimensions included in them. Therefore, both indexes have also been subject to extensive revision and have already undergone various methodological changes.

4.1 Foundations of the composite indicators

Theoretical framework: the three composites, even when not explicitly including the word "development" in its name, try to capture a development concept in its multidimensionality. The authors of the three indices propose a definition informed both by literature and expert opinion. However, the UNDP is the only one that consistently accompanies the definition of human development for the HDI of its theoretical underpinnings: by quoting Amartya Sen's capabilities approach.

In the case of the LPI and the SPI conceptual groundings, their authors designed the corresponding definitions of prosperity and social progress. Evidently, both definitions were constructed based on underlying theory, which is apparent while revising the references to literature that are made in the documentation through which the indexes have been disclosed since their creation.

[24] In this study, accessibility is referred to the availability of the information (e.g. explanations, data, among others) in the most recently published index methodology's documentation. Accessibility to the information may also be improved by promptly referring the reader to the document that contains the additional explanations or data.

The LPI authors thoroughly discuss their definition of prosperity, along with each domain, pillar, and element description, in every year's documentation. On the other hand, though in the 2014-2017 SPI's reports the authors acknowledged that Sen's capabilities approach inspires the social progress concept, this recognition is not apparent in the 2020 SPI's documentation: dimensions and components are very briefly described without references to literature. Although the degree of accessibility may vary, it is evident that the three indexes are grounded in literature and expert opinions, which adds to the credibility of these indexes.

Regarding the composite's objective, both the LPI and SPI are intended to measure the proposed concept and provide a benchmark of success in this regard. The LPI's country profiles and the SPI's country-specific scorecards disclose each country's position in the global ranking per component: from each underlying indicator until the overall index. By providing this series of scores and rankings, the reader can identify at a glance the potential drivers of overall performance in prosperity or social progress according to each index framework. On the other hand, the HDI has the objective to be a summary measure of achievements; thus, it seems fitting that it is regularly disclosed with a collection of other indexes that complement it.

Data selection and imputation of missing data: while revising each index's components and their underlying indicators, it could be argued that they share some core similarities. As the three indexes try to measure related concepts, they also include similar underlying indicators in their calculations. For example, "life expectancy at birth" is used by the HDI, while "life expectancy at 60" is present in the LPI and the SPI. Some measure of school enrolment is used in the three indexes, though with different demographic focus. Income was deliberatively left out of the SPI given its explicit non-economic focus; LPI uses the GDP per capita growth rate while HDI uses the GNI in its logarithmic form.

Each time there is a change in HDI's indicators, they accompany it with justifications, providing a detailed account of the reasons that motivate their inclusion in the index. Additionally, data limitations, either in country coverage or differences in quality, are quoted as the main reasons to maintain the current four indicators (Klugman et al., 2011).

Both the Social Progress Imperative and the Legatum Institute provide a set of criteria that guide the selection of indicators. Understandably, the high amount of indicators that compose the SPI and the LPI would render it complicated to provide a detailed account of the sufficiency to which each indicator fulfills the criteria; an even further step would be to discuss all the alternative indicators that were discarded on the same grounds. Nevertheless,

LPI's and SPI's reports include a section each year explaining changes in indicators, either as a replacement or a newly measured aspect.

The raw dataset (all data gathered directly from the stated sources) for each of the three composites are freely available for download. Either on the data files (.xlsx) or the methodological reports, the authors state for each indicator the description, sources, units, and period to which the data corresponds. However, the final dataset, including all imputed values, is only available for the HDI[25]. If the final datasets were fully disclosed for both the LPI and SPI, the composite's calculation could be autonomously replicated by interested readers and allow the external audience to conduct complete independent studies of the index's properties.

In the case of the LPI, as mentioned, there is not a complete account of the imputed values and the strategy used in each case, which is particularly relevant due to the high reliance on imputations that this composite exhibits. Even when they disclose the percentage of imputed values, and that they affirm that data imputation -as well as reliability of the sources- are reasons for down-weighting the indicators, there is no discussion of how the high amount of indicators imputed -and the techniques employed- may affect the overall index results.

For the SPI, within the disclosed dataset, they provide those values imputed before calculating the index, along with the technique applied in each case; however, as they comment on their methodological report, other imputations are performed during the SPI's calculation process that are not revealed.

Data transformations are reported on a case-by-case basis for each index. Nevertheless, a discussion of the implications of the performed transformation is only provided for the HDI log transformation of GNI per capita. As discussed before, the log transformation of the income component is part of a broader critique related to the implicit trade-offs of the index; however, the UNDP proposes a theoretical explanation to continue applying this transformation of the income variable. On the other hand, data transformations for the LPI and SPI are performed for scaling purposes or for treating outliers; as mentioned, though all transformed indicators are reported, the possible effect of the transformation on the index results is not disclosed for these two composite indicators.

[25] However, the regressions and formulas applied for estimation of the imputed values of the HDI is not available.

4.2 Structural form of the composite indicators

Multivariate analysis: multivariate analysis has been performed for the three indices. In the case of the SPI and the LPI, the authors estimate the Cronbach's alpha coefficient and follow an established standard of 0.7 or above as proof of the underlying indicators being jointly relevant to explain the component to which they belong; however, there is no mention of indicators that may have been discarded for exhibiting too low correlation within a prospective grouping.

In the case of the SPI, the coefficients at the component level are revealed in the methodological report. Additionally, the case of the only component that has a coefficient lower than 0.7 is commented and justified on the grounds of a lack of alternative indicators. Additionally, though a complete account of correlations between SPI's underlying indicators is not revealed, the authors disclose the range deemed acceptable for correlations.

Conversely, for the LPI only the coefficients at the pillar level are disclosed, with two of them being lower than 0.7. The Institute does not reveal the coefficients of any element, and only recognizes that a non-disclosed number of them are below 0.7. Despite the statistical evidence against their inclusion, no justification for it is provided. Only a general comment is recorded on the documentation: they were kept on the grounds of perceived conceptual importance.

In the case of the HDI, correlation analysis has been performed in previous years, but it is not yearly published; this analysis must be traced back to previous editions of the HDR or to complementary documents published by the HDRO[26]. No evidence has been found that the Cronbach's alpha coefficient was calculated for the current set of indicators of the HDI. Additionally, while analyzing the correlations between indicators, it is important to consider that a high correlation may signal redundancy among the index components, which would make it irrelevant to aggregate them in a composite. As mentioned before, the HDI dimensions are indeed highly correlated; although the HDRO recognizes the issue of possible redundancy, the critique is dismissed by referring to correlation analysis of the changes over time of its components and the HDI scores. Nevertheless, the statistical problem should be further commented on and recognized in the latest methodological documentation.

[26] As mentioned, a discussion of the relation between the index variables is present in Klugman et al, (2011), Kovacevic (2011) and García and Kovacevic (2011).

Normalization: the min-max formula is employed in the three indices[27]. The reasons for favoring this method of normalization instead of others are not disclosed in the methodological reports. However, in the case of the HDI, García and Kovacevic (2011) comment how applying z-scores would not fit the HDI framework, as it would allow some degree of compensability by rewarding exceptional performance in one of the indicators. In the case of the LPI and SPI, no discussion is provided for favoring this method. García and Kovacevic (2011) mention the easiness of interpretation of an index ranging from 0 to 100 as a convenient feature of the min-max approach, which may be why it is applied in the three composite indicators.

Additionally, as a necessary step, SPI's authors utilize z-scores to normalize indicator values prior to generating weights through PCA; therefore, this method is justified as part of a statistical process, and the authors do not further comment on the effect of this strategy within the index theoretical framework.

The setting of best- and worst-case scenarios are performed for each index. For the HDI, these bounds are justified for each indicator, either referring to relevant literature, historical values, or to a conceptual meaning in terms of capabilities (e.g., societies can live without education; therefore, the lower bound is set at 0). On the other hand, both the SPI and LPI only describe in general terms how those scenarios are generated (i.e. they mention that they correspond mainly to logical or theoretical bounds). Though these explanations may be enough to understand some cases, it would be more transparent to provide a rationale for setting bounds for each indicator, preferably by referring to the adopted theoretical framework (i.e., the bounds fit with achieving the highest or lowest level of prosperity or social progress). Moreover, for the LPI, the values of the best and worst-case scenarios are not revealed, so it is not possible to verify the normalization process.

Weighting: all nominal weights are revealed for the three indexes. However, the process and justification for determining them are different in each case. The UNDP assigns equal weights for each of HDI's components based on its theoretical framework: capabilities captured in each dimension cannot be deemed relatively more important than the others in contributing to human development. However, as mentioned, García and Kovacevic (2011) recognize that the importance (implicit weights) of each dimension on the overall index is

[27] As mentioned in the previous Chapter, the Legatum Institute applies the DTF method, which is used the same formula as a min-max approach with fixed goalposts.

also affected by other aspects of the construction process, such as data properties and normalization.

The theoretical justification for the assigned nominal weights is not explicit for the LPI and SPI. For the LPI, experts assign weights for indicators and elements based on a set of criteria and statistical reasoning; but the rationale is not explained for each of the cases. When considering that the data quality is one of the criteria to down-weight indicators, it would be misleading to portray weights as a measure of relative importance, as intended by the Legatum Institute. Subsequently, no further explanation is offered for favoring equal weighting on pillars and domains.

In the case of the SPI, weights are generated through PCA at the indicator level. Even when the software and command to execute the statistical method are revealed, since the final dataset with all imputed values is not provided, it is impossible to replicate the process independently. Additionally, though PCA is a statistical process that focuses on avoiding double-counting, and has associated statistical advantages and shortcomings, applying the process without commenting the resulting weights omits a normative consideration of what these distinct indicator's weights mean in contributing to social progress. SPI's authors assign equal weights to components and dimensions; rather than providing theoretical reasons to support the validity of equal weighting at both stages, they seemingly favored them after an elimination of alternatives: after testing other statistical approaches, the authors mentioned that there was not not a sufficiently robust relative relationship, both theoretically and empirically, to favor any other weighting scheme; however, the last time they disclosed this analysis was on 2017, and under a different collection of underlying indicators.

Additionally, as mentioned on Chapter 2, the research conducted by Paruolo et al. (2013), Becket et al. (2017), and Schlossarek et al. (2019) have found that the importance differs from the nominal weights and is affected by aspects such as the correlation between variables, transformations, normalization, and aggregation method. Therefore, authors should recognize that the importance of each component of an index is determined by far more than the nominal weight set. Although the UNDP has recognized that equal weights are not a reflection of the actual effect of each dimension on the index's result, this explanation is not immediately accessible, as it has to be traced to documents published more than ten years ago.

Aggregation: the aggregation methods are disclosed for each index. However, only for the HDI is a discussion provided on the implications of using a geometric aggregation, which is still a reason for debate, as recognized in the 2020 HDR. Particularly, justifications were warranted due to the change in 2010 from the arithmetic mean to geometric mean to aggregate the dimensions of the HDI, and they were sustained by referring to the theoretical framework: compensability should not be allowed as capabilities cannot be traded one for another.

As in other cases, the scrutiny faced by the HDI prompted a response by the HDRO. They explicitly commented that trade-offs (MRS) between dimensions of the HDI are relevant, but dismissed the critiques by resorting to explanations grounded in the capabilities approach, that is, claiming that the HDI does not propose a social welfare function that must be maximized (Klugman et al., 2011). Even if critiques have been dismissed and trade-offs have not been computed, this aspect was mentioned by Klugman et al. (2011) and on the 2020 HDR.

Both the SPI and LPI aggregate firstly as weighted averages on the lower levels of the index -as discussed, they assign different weights to underlying indicators-, and later employ arithmetic mean to generate both dimension's and index' scores. Even when this method allows for compensability between equally weighted dimensions, the authors do not provide a discussion about this issue and its implications (e.g., on a country whose indicators present heterogeneous levels of achievements across dimensions).

4.3 Robustness tests

Uncertainty and sensitivity analysis: the authors of the three indices have carried out robustness tests; however, the number of decisions tested, and alternative scenarios considered, is variable. Understandably, it would be a significant effort to test alternatives to every decision made in the construction process. Nevertheless, it is advisable to consider how the index's results would change if alternative choices were made at the steps that crucially define the functional form of the index: normalization, weighting, and aggregation.

For the SPI, when they ran robustness tests in 2017, and for the LPI in 2019 and 2020, the authors mainly test the weighting choices in various scenarios[28]. Though authors of both indices conclude that the index is robust due to the low number of ranking variations in the alternative scenarios, this also opens to debate if the chosen weights provide any value added

[28] Externally, the JRC did evaluate the aggregation method of the SPI through geometric mean in 2018.

to the measure. However, neither of them discloses the ranking variations for the entire country sample, although they do discuss what they deem as the most salient results (i.e., those cases in which a country ranking is significatively altered).

Many of the decisions involved in the HDI construction process are discussed or statistically tested by García and Kovacevic (2011). They often offer some justification by recalling the theoretical framework; these justifications have been referred to in the previous Chapter when appropriate. For the decisions that they submit to sensitivity analysis, they develop a detailed account of the statistical process and of the results obtained, both narratively and through graphical visualization. However, this type of analysis is not regularly performed in the case of the HDI. Understandably, if no methodological changes were introduced in the index's structure, it would not be entirely necessary to perform this extensive analysis again. Nevertheless, the UNDP may update such analysis and ensure the accessibility to it by referencing it in the Technical Note of the respective HDR.

4.4 Analysis of the composite indicator's results against their components and related indicators

Back to the real data and links to other indicators (or variables): the UNDP has contrasted the results of the HDI to its components and other indicators for the 2009 and 2010 versions of the index; however, this exam is not performed and disclosed yearly. Given the high number of underlying indicators for both the SPI and LPI, it is understandable that no systematic correlation or causality analysis has been disclosed between the index results and each of the data series that compose it. However, even when the results are not commented, the LPI reports include the Pearson correlation coefficients between the pillars and the overall index score. Furthermore, the Legatum Institute presents a correlation analysis between LPI's results and other composite indicators, commenting on its similarities and differences. In the case of the SPI, this type of analysis has not been found on the 2020 methodological report.

From the discussion, it is possible to derive recommendations to improve the transparency with which the methodological aspects of each index are disclosed; these are summarized in Table 4.

Table 4 Recommendations for improving transparency of composite indicators of development

Section	Step	Recommendations
Foundations of the composite indicator: theory and variables	Theoretical framework	- Improve accessibility* to the description of the components of the index (SPI)
	Data selection	- Make criteria for data selection clearer (SPI, LPI) - Reveal and justify which indicators were included even when criteria were not fulfilled (e.g., point out lack of other alternatives) (LPI) - Discuss the possible effect of transformations (SPI, LPI)
	Imputation of missing data	- Provide a complete dataset including all imputed values and the method to obtain them (SPI, LPI; only method: HDI) - Explain the effect of imputed values on the index score (due to high reliance on imputations: LPI)
Structural form of the index	Multivariate analysis	- Reveal all the test results (for each level of the index) (LPI, HDI) - Present and discuss the analysis for the most recent version of the index (HDI)
	Normalization	- Reveal all the relevant inputs for applying the normalization method: best- and worst-case scenarios (LPI) - Reveal the rationale for setting best and worst-case scenarios per each indicator (SPI, LPI)
	Weighting and aggregation	- Make the rationale for setting weights clearer (LPI) - Discuss the assigned weights within the adopted theoretical framework (SPI, LPI) - Recognize the difference between the nominal weights set and the importance of each component in the index results (LPI, SPI; improve accessibility*: HDI) - Discuss the allowed compensability -or not- between index dimensions and its implications according to the adopted theoretical framework (SPI, LPI; improve the accessibility* of discussion: HDI) - Disclose and comment the trade-offs between indicators, that are affected by the chosen weights, aggregation method, and data properties and its transformations (LPI, SPI; disclose the trade-offs and improve accessibility* to theoretical explanation: HDI)
Robustness tests	Uncertainty and sensitivity analysis	- Conduct sensitivity analysis for the most recent methodology of the index (HDI, SPI) - If some decision (step) is left out of the analysis, justify it (SPI, LPI; provide it for the most recent methodology: HDI) - Reveal the complete account of results of the analysis (e.g., the complete list of ranking variations) (LPI, SPI, HDI)
Analysis of the index results against its components and related indicators	Back to the real data	- Perform and reveal analysis for the most recent index's methodology (SPI, HDI; provide a comment on the resulting coefficients: LPI)
	Links to other indicators (variables)	- Perform and reveal the analysis for the latest index's methodology (HDI, SPI)

*Improve accessibility: it is suggested to mention the referred information or analysis on the documentation where the methodology is disclosed each year or provide a clear indication of where such information or analysis can be found (preferably enclosed on the methodological report).
Source: author.

Conclusions

Composite indicators summarize complex concepts in a single measure. Notably, composites that rank countries' performance are powerful tools in the development field, as they induce debate at the national and international levels, potentially having policy implications. Therefore, authors should be clear in communicating the theory and methodology employed to aid in the proper interpretation and use of the measure.

The index's construction process involves many steps, where various decisions must be made. Those decisions, their results and implications should be transparently disclosed to the public. Being transparent does not diminish the subjectivity of the decisions: however, a transparent disclosure of the choices does contribute to making them accountable. Moreover, the exposure to constructive criticism gives the opportunity of refining the methodology and improve the measurement of the proposed concept.

Though the three composite indicators studied -SPI, HDI and LPI- differ in theoretical and methodological grounds, their joint analysis sheds light on aspects where transparency could be improved to aid the reader in understanding the index's properties and limitations. Therefore, the study advocates for improving transparency and offers generally applicable recommendations in this regard.

Concerning the theoretical framework, the three indexes disclose a definition for the concept they try to capture. Dimension's and subcomponents' descriptions are sustained on theory and empirical research. However, these definitions and their references to underlying theory sometimes have to be traced to previous documents. Therefore, it is advisable to retain the description of each component in every year's documentation to facilitate understanding the procedure within the theoretical framework.

The selection criteria for underlying indicators are disclosed for the SPI and the LPI; however, these criteria could be more explicit. In some cases, disclosing the rationale for including each indicator may be necessary to understand why it was favored instead of other alternatives. The raw data used for calculating each index is published in .xlsx files. However, the final datasets, including all the imputed values, are not disclosed for the SPI and LPI. Providing the complete dataset would allow for independent reviews. Additionally, in the case of the LPI, the discussion of the effect of imputations may be necessary since the index relies heavily on data imputation, so the readers should be careful in the conclusion they

drawn from the results. Therefore, a transparent recognition of the effect that imputed values may have on composites that rely heavily on this practice is recommended.

Normalization, weighting, and aggregation are fundamental steps that define the structure of the composite indicator. The study has found that authors of the three composites reveal the strategy used for each step; however, justifications and a discussion of their implications are absent in many cases.

Weights are defined through distinct processes. HDI and LPI weights are primarily set by referencing the theoretical framework or relying on expert opinion, but the rationale is not disclosed on an indicator by indicator basis for the LPI. Conversely, for the SPI, underlying indicators are weighted through a statistical method (PCA), the process is thoroughly described and intends to avoid issues of double counting; however, the different weights generated are not explained within the theoretical framework

Therefore, as a general recommendation, the authors should provide the rationale for weighting within the adopted conceptual framework. Additionally, they should explicitly recognize that the importance of each component in the overall index's result is not only affected by nominal weights but also by data transformations, correlations, normalization and aggregation method.

The criticism received by the HDI has prompted the UNDP to discuss the compensability between the HDI's dimensions, and to use the geometric mean for aggregation. On the other hand, the LPI and SPI use the arithmetic mean without discussing the compensability allowed between dimensions. Therefore, it is recommended that authors duly mention how both aggregation and weighting decisions influence the compensability and the implied trade-offs between the index dimensions, as these considerations are essential in understanding the measured concept through a composite indicator.

Robustness checks in the form of sensitivity analysis have been performed for the three indices. However, LPI and SPI analysis results are not entirely disclosed, and just extreme cases are discussed. Disclosing the full list of countries' ranking variations would be helpful to assess the performance of countries not discussed by the authors. Additionally, it is recommended to test other choices, such as aggregation method, and to update the analysis when methodological changes are introduced.

It is useful that the authors disclose correlation analysis between index components, as these analyses aid in understanding the interplay between dimensions and the result it will have in

the index. In general, multivariate analysis, the study of the correlation between the index and its components -and other variables- can shed light on which are the index drivers and on its value-added.

Overall, a call for improving the accessibility to the core rationale, considerations, and implications of the index's methodological process and properties is made. Although analyses and considerations can sometimes be traced back to earlier documents, it may be difficult for the interested public to find them. Without necessarily reproducing the complete information on each year's documentation, a list could be made available to guide the reader to the source where further explanations can be found in each case. As stated before, transparency is fundamental while presenting a composite indicator's construction process: it is crucial to provide enough information to describe the relationship and interplay that the index allows between the dimensions of the proposed development concept.

References

Anand, S. (2018). Recasting Human Development Measures. *Discussion Paper*. UNDP Human Development Report. http://hdr.undp.org/sites/default/files/anand_recasting_human_development_measures.pdf

Bandura, R. (2008). A Survey of Composite Indices Measuring Country Performance. *UNDP/ODS Working Paper*. UNDP. https://www.undp.org/content/dam/undp/library/corporate/Development%20Studies/indices_2008_bandura.pdf

Barrington-Leigh, C., & Escande, A. (2018). Measuring Progress and Well-Being: A Comparative Review of Indicators. *Social Indicators Research*, *135*(3), 893–925. https://doi.org/10.1007/s11205-016-1505-0

Becker, W., Benavente, D., Dominguez-Torreiro, M., Moura, C., Neves, A., Saisana, M., & Vertesy, D. (2019). *COIN tool: User guide*. Publications Office of the European Union. https://data.europa.eu/doi/10.2760/523877

Becker, W., Saisana, M., Paruolo, P., & Vandecasteele, I. (2017). Weights and importance in composite indicators: Closing the gap. *Ecological Indicators*, *80*, 12–22. https://doi.org/10.1016/j.ecolind.2017.03.056

Decancq, K., & Lugo, M. A. (2009). Setting Weights in Multidimensional Indices of Well-being and Deprivation. *Working Paper* 18. Oxford Poverty & Human Development Initiative (OPHI). https://www.ophi.org.uk/wp-content/uploads/OPHI-wp18.pdf

Freudenberg, M. (2003). *Composite Indicators of Country Performance: A Critical Assessment* (OECD Science, Technology and Industry Working Papers No. 2003/16; OECD Science, Technology and Industry Working Papers, Vol. 2003/16). https://doi.org/10.1787/405566708255

García, C., & Kovacevic, M. (2011). Uncertainty and Sensitivity Analysis of the Human Development Index. *Human Development Research Paper* 47. UNDP. http://hdr.undp.org/sites/default/files/hdrp_2010_47.pdf

Gidwitz, Z., Heger, M. P., Pineda, J., & Rodríguez, F. (2010). Understanding Performance in Human Development: A Cross-National Study. *Human Development Research Paper* 42. UNDP. http://hdr.undp.org/sites/default/files/hdrp_2010_42.pdf

Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2019). On the Methodological Framework of Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness. *Social Indicators Research*, *141*(1), 61–94. https://doi.org/10.1007/s11205-017-1832-9

Kahneman, D., & Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences*, *107*(38), 16489–16493. https://doi.org/10.1073/pnas.1011492107

Killingsworth, M. A. (2021). Experienced well-being rises with income, even above $75,000 per year. *Proceedings of the National Academy of Sciences*, *118*(4), e2016976118. https://doi.org/10.1073/pnas.2016976118

Klasen, S. (2018). Human Development Indices and Indicators: A Critical Evaluation. *Background Paper, Human Development Report Office*. UNDP. http://hdr.undp.org/sites/default/files/klasen_final.pdf

Klugman, J., Rodríguez, F., & Choi, H.-J. (2011). The HDI 2010: New controversies, old critiques. *The Journal of Economic Inequality*, *9*(2), 249–288. https://doi.org/10.1007/s10888-011-9178-z

Kovacevic, M. (2011). Review of HDI Critiques and Potential Improvements. *Human Development Research Paper* 33. UNDP. http://hdr.undp.org/en/content/review-hdi-critiques-and-potential-improvements

Legatum Institute. (2020a). The 2020 Legatum Prosperity Index: A tool for transformation. https://docs.prosperity.com/2916/0568/0539/The_Legatum_Prosperity_Index_2020.pdf

Legatum Institute. (2020b). Defining Prosperity. https://docs.prosperity.com/6516/0586/8116/Defining_Prosperity.pdf

Legatum Institute. (2020c). Measuring Prosperity. https://docs.prosperity.com/4116/0586/8547/Measuring_Prosperity.pdf

Legatum Institute. (2020d). *Our Advisors*. Legatum Prosperity Index Website. https://www.prosperity.com/about/our-advisors

Legatum Institute. (2020e). Legatum Prosperity Index: Full 2020 Data Set. https://www.prosperity.com/download_file/view/4062/2025

Legatum Institute. (2020f). The Legatum Prosperity Index 2020: Country Profiles. https://docs.prosperity.com/1916/0568/0974/Legatum_Prosperity_Index_2020_Country_profiles.pdf

Legatum Institute. (2020g). *About: Methodology*. Legatum Prosperity Index Website. https://www.prosperity.com/about/methodology

Legatum Institute. (2019a). The 2019 Legatum Prosperity Index: A tool for transformation. https://www.prosperity.com/download_file/view_inline/3924

Legatum Institute. (2019b). The 2019 Legatum Prosperity Index Methodology Report. https://docs.prosperity.com/7515/8634/9002/Methodology_for_Legatum_Prosperity_Index_2019.pdf

Legatum Institute. (2018). The 2018 Legatum Prosperity Index: Methodology Report. https://prosperitysite.s3-accelerate.amazonaws.com/7915/4398/9989/20181204_LI_PI_Methodology_Report.pdf

Legatum Institute. (2017). The 2017 Legatum Prosperity Index: Methodology Report. http://prosperitysite.s3-accelerate.amazonaws.com/9615/1186/6075/Legatum_Prosperity_Index_2017_Methodology_Report.pdf

Legatum Institute. (2016). The 2016 Legatum Prosperity Index: Methodology Report. https://www.prosperity.com/download_file/view_inline/3924

Lustig, N. (2011). Multidimensional indices of achievements and poverty: What do we gain and what do we lose? An introduction to JOEI Forum on multidimensional poverty. *The Journal of Economic Inequality*, *9*(2), 227–234. https://doi.org/10.1007/s10888-011-9186-z

Munda, G., & Nardo, M. (2009). Noncompensatory/nonlinear composite indicators for ranking countries: A defensible setting. *Applied Economics*, *41*(12), 1513–1523. https://doi.org/10.1080/00036840601019364

Norlén, H., & Caperna, H. (2018). *The JRC statistical audit of the Social Progress Index (SPI).* Publications Office of the European Union. https://data.europa.eu/doi/10.2760/343522

OECD. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. OECD Publications.

Paruolo, P., Saisana, M., & Saltelli, A. (2013). Ratings and rankings: Voodoo or science? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *176*(3), 609–634. https://doi.org/10.1111/j.1467-985X.2012.01059.x

Porter, M., Scott, S., & Green, M. (2017). Social Progress Index 2017. https://www.socialprogress.org/static/20d068d269a282c3edfd41e1a1b3b7d2/2017-social-progress-index.pdf

Ravallion, M. (2012a). Mashup Indices of Development. *The World Bank Research Observer*, *27*(1), 1–32. https://doi.org/10.1093/wbro/lkr009

Ravallion, M. (2012b). Troubling tradeoffs in the Human Development Index. *Journal of Development Economics*, *99*(2), 201–209. https://doi.org/10.1016/j.jdeveco.2012.01.003

Ravallion, M. (2011). The human development index: A response to Klugman, Rodriguez and Choi. *The Journal of Economic Inequality*, *9*(3), 475–478. https://doi.org/10.1007/s10888-011-9193-0

Saisana, M., Saltelli, A., & Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *168*(2), 307–323. https://doi.org/10.1111/j.1467-985X.2005.00350.x

Saltelli, A. (2007). Composite Indicators between Analysis and Advocacy. *Social Indicators Research*, *81*(1), 65–77. https://doi.org/10.1007/s11205-006-0024-9

Schlossarek, M., Syrovátka, M., & Vencálek, O. (2019). The Importance of Variables in Composite Indices: A Contribution to the Methodology and Application to Development Indices. *Social Indicators Research*, *145*(3), 1125–1160. https://doi.org/10.1007/s11205-019-02125-9

Sébastien, L., & Bauler, T. (2013). Use and influence of composite indicators for sustainable development at the EU-level. *Ecological Indicators*, *35*, 3–12. https://doi.org/10.1016/j.ecolind.2013.04.014

Seth, S., & McGillivray, M. (2018). Composite indices, alternative weights, and comparison robustness. *Social Choice and Welfare*, *51*(4), 657–679. https://doi.org/10.1007/s00355-018-1132-6

Sirgy, M. J. (2011). Theoretical Perspectives Guiding QOL Indicator Projects. *Social Indicators Research*, *103*(1), 1–22. https://doi.org/10.1007/s11205-010-9692-6

Social Progress Imperative. (2020a). Social Progress Index dataset 2011-2020. https://legacy.socialprogress.org/assets/downloads/2011-2020-Social-Progress-Index.xlsx

Social Progress Imperative. (2020b). Social Progress Index 2020 Scorecard: El Salvador. https://www.socialprogress.org/static/36ef49db327f6233ff01b757cca95844/slv-scorecard.pdf

Stern, S., Krylova, P., & Harmacek, J. (2020). 2020 Social Progress Index Methodology Summary. https://www.socialprogress.org/static/1aa2d19690906eb93c6cdb281e5ee68b/2020-social-progress-index-methodology.pdf

Stern, S., Wares, A., & Epner, T. (2018). Social Progress Index 2018 Methodology Summary. https://www.socialprogress.org/static/9b25a2a187a2638e806fd39d0a7e3ab0/2018-social-progress-index-methodology.pdf

Stern, S., Wares, A., & Epner, T. (2017). Social Progress Index 2017 Methodology Report. https://www.socialprogress.org/static/875b9facd696ca43a16db16edd9031be/2017-social-progress-index-methodology.pdf

Stern, S., Wares, A., & Hellman, T. (2016). Social Progress Index 2016 Methodology Report. https://www.socialprogress.org/static/b065a359a285d66e892798f79d03f7f6/2016-social-progress-index-methodology.pdf

Stern, S., Wares, A., & Orzell, S. (2015). Social Progress Index 2015 Methodology Report. https://www.socialprogress.org/static/5202049f62cb773e9dd27a798f1633c8/2015-social-progress-index-methodology.pdf

Stern, S., Wares, A., Orzell, S., & O'Sullivan, P. (2014). Social Progress Index 2014 Methodology Summary. https://www.socialprogress.org/static/82a7f907b051ef1cd040c580ac5f497d/2014-social-progress-index-methodology.pdf

Stiglitz, J., Sen, A., & Fitoussi, J.-P. (2009). Report by the Commission on the Measurement of Economic Performance and Social Progress. https://ec.europa.eu/eurostat/documents/8131721/8131772/Stiglitz-Sen-Fitoussi-Commission-report.pdf

UNDP. (1990). Human development report 1990. Oxford University Press. http://hdr.undp.org/en/reports/global/hdr1990

UNDP. (1994). Human development report 1994. Oxford University Press. http://www.hdr.undp.org/en/content/human-development-report-1994

UNDP. (2010). Human Development Report 2010: The real wealth of nations: Pathways to human development. http://hdr.undp.org/en/content/human-development-report-2010

UNDP. (2011). Human Development Report 2011 Sustainability and equity: A better future for all. http://hdr.undp.org/en/content/human-development-report-2011

UNDP. (2013). Human Development Report 2013: Technical notes. http://hdr.undp.org/sites/default/files/hdr_2013_en_technotes.pdf

UNDP. (2020a). Human Development Report 2020: The next frontier human development and the anthropocene. http://hdr.undp.org/sites/default/files/hdr2020.pdf

UNDP. (2020b). Human Development Report 2020: Technical notes. http://hdr.undp.org/sites/default/files/hdr2020_technical_notes.pdf

UNDP. (2020c). Human Development Index Statistical Annex. http://hdr.undp.org/sites/default/files/2020_statistical_annex_all.xlsx

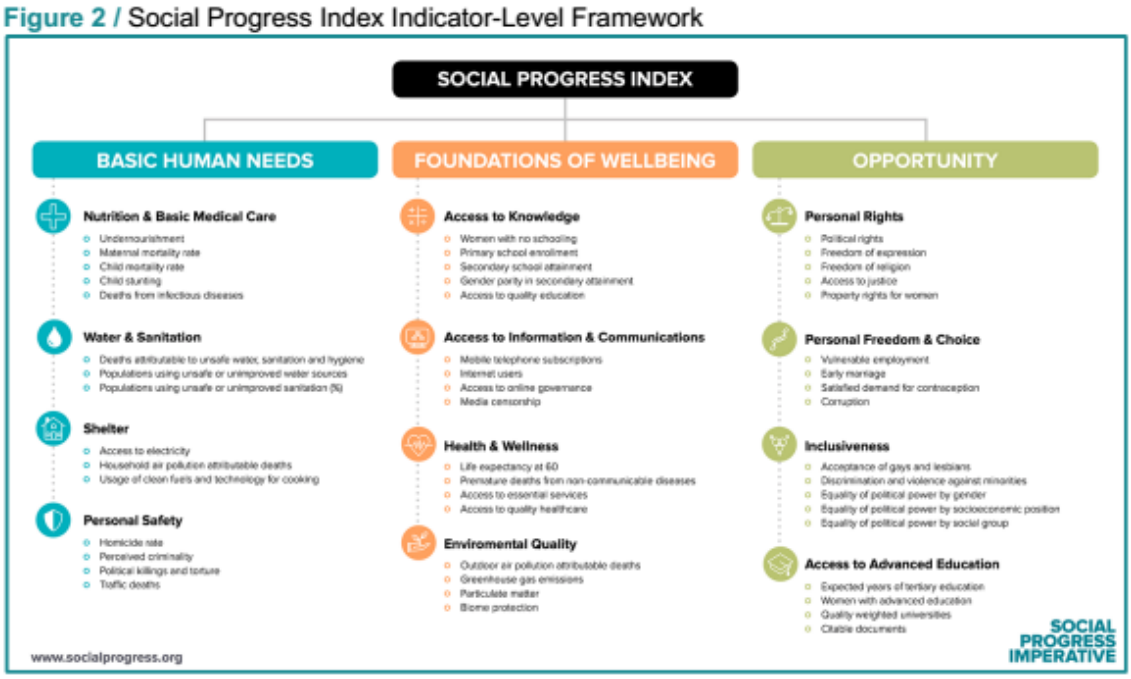
UNECE. (2019). Guidelines on producing leading, composite and sentiment indicators. http://www.unece.org/fileadmin/DAM/stats/publications/2019/ECECESSTAT20192.pdf

Yang, L. (2014). An Inventory of Composite Measures of Human Progress. *Occasional Paper on Methodology*. UNDP. http://hdr.undp.org/sites/default/files/inventory_report_working_paper.pdf

Appendices

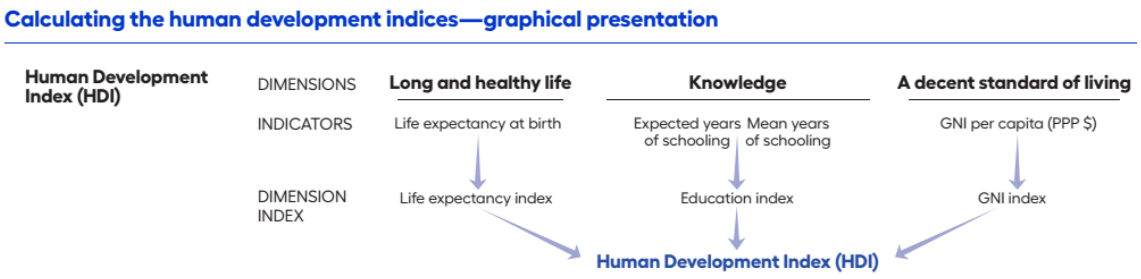
Appendix A: structure of the composite indicators.

Figure 1: Social Progress Index' structure



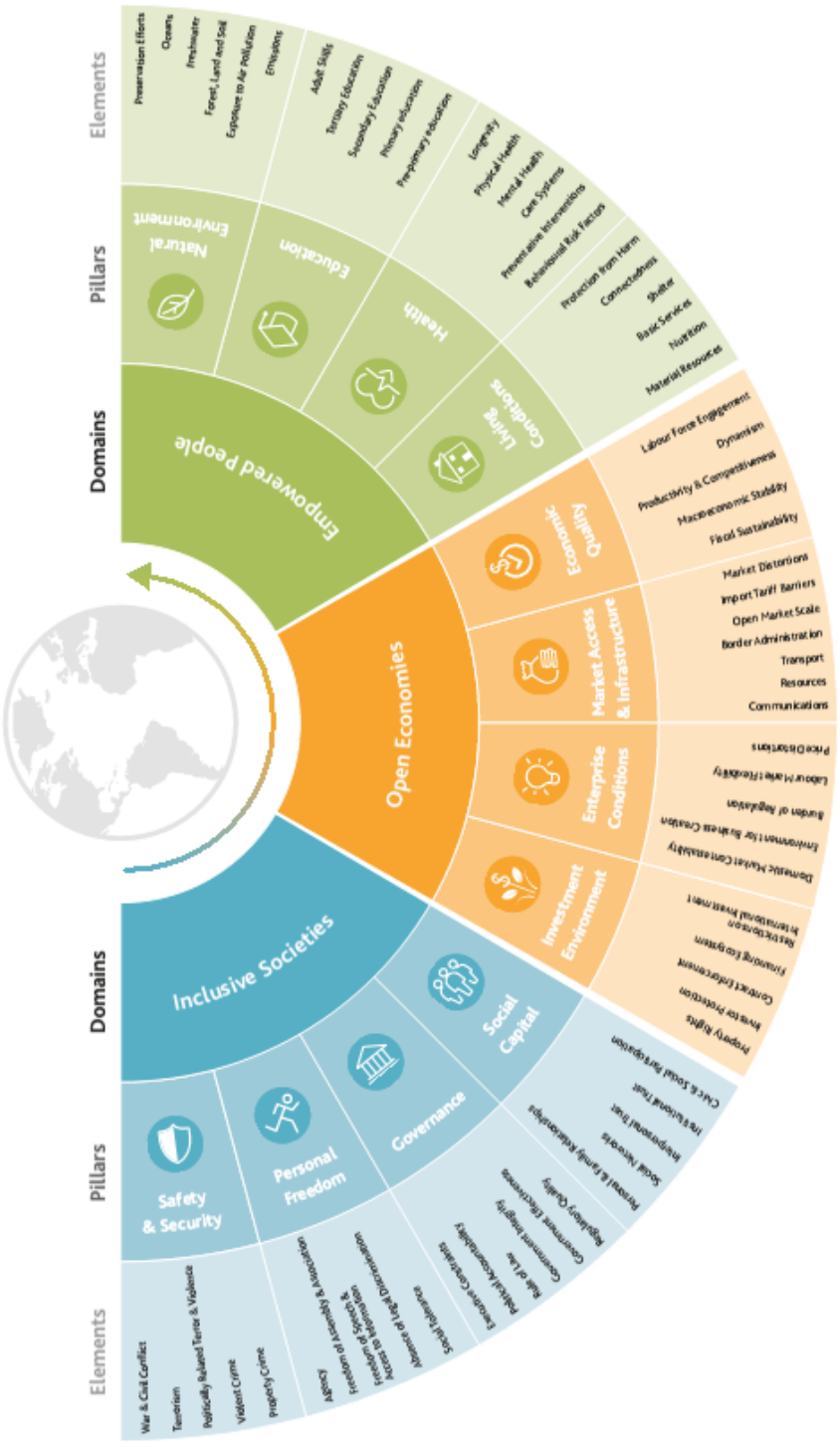
Source: Stern et al., 2020, p. 7.

Figure 2: Human Development Index' structure



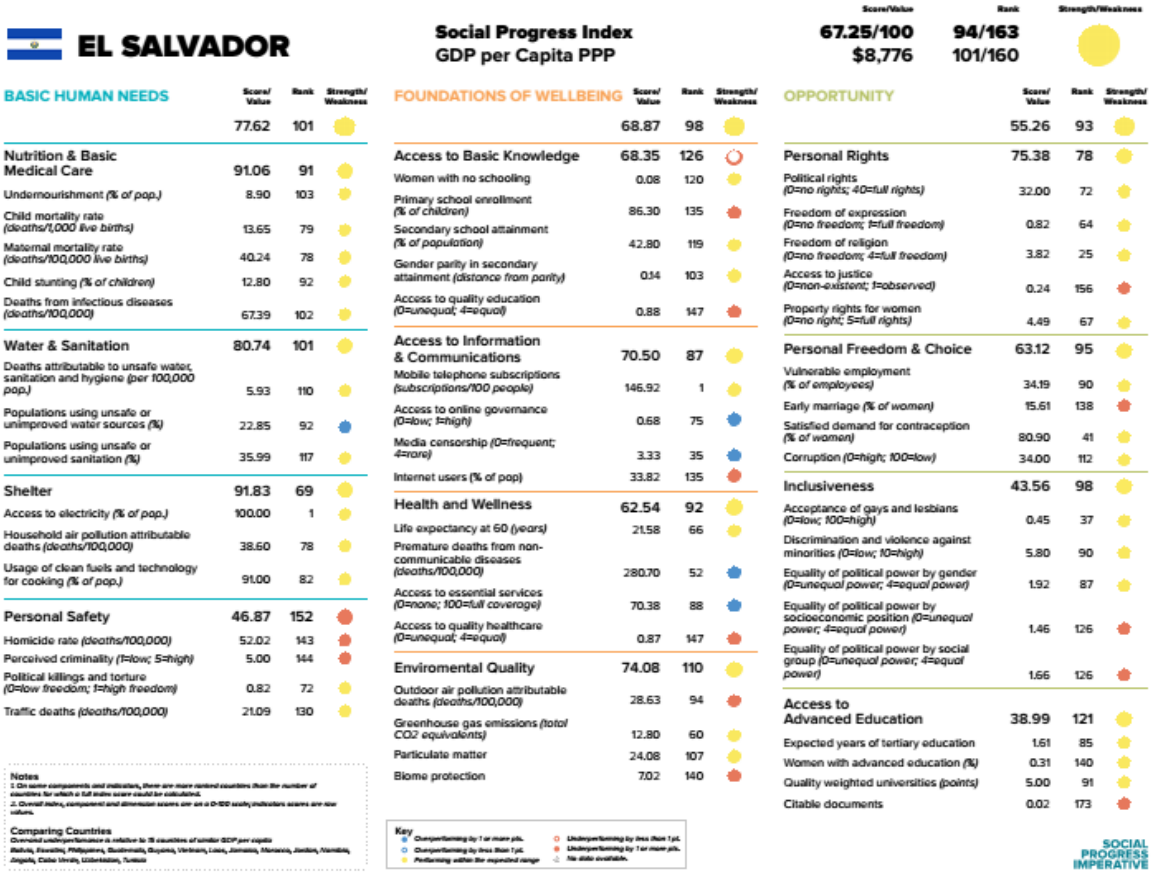
Source: UNDP, 2020b, p. 1.

Figure 3: Legatum Prosperity Index's structure



Source: Legatum Institute, 2020a, p. 35.

Appendix B: Social Progress Index scorecard example



EL SALVADOR

Social Progress Index
GDP per Capita PPP

	Score/Value	Rank
	67.25/100	**94/163**
	$8,776	**101/160**

BASIC HUMAN NEEDS	Score/Value	Rank	Strength/Weakness
	77.62	101	
Nutrition & Basic Medical Care	91.06	91	
Undernourishment *(% of pop.)*	8.90	103	
Child mortality rate *(deaths/1,000 live births)*	13.65	79	
Maternal mortality rate *(deaths/100,000 live births)*	40.24	78	
Child stunting *(% of children)*	12.80	92	
Deaths from infectious diseases *(deaths/100,000)*	67.39	102	
Water & Sanitation	80.74	101	
Deaths attributable to unsafe water, sanitation and hygiene *(per 100,000 pop.)*	5.93	110	
Populations using unsafe or unimproved water sources *(%)*	22.85	92	
Populations using unsafe or unimproved sanitation *(%)*	35.99	117	
Shelter	91.83	69	
Access to electricity *(% of pop.)*	100.00	1	
Household air pollution attributable deaths *(deaths/100,000)*	38.60	78	
Usage of clean fuels and technology for cooking *(% of pop.)*	91.00	82	
Personal Safety	46.87	152	
Homicide rate *(deaths/100,000)*	52.02	143	
Perceived criminality *(1=low; 5=high)*	5.00	144	
Political killings and torture *(0=low freedom; 1=high freedom)*	0.82	72	
Traffic deaths *(deaths/100,000)*	21.09	130	

FOUNDATIONS OF WELLBEING	Score/Value	Rank	Strength/Weakness
	68.87	98	
Access to Basic Knowledge	68.35	126	
Women with no schooling	0.08	120	
Primary school enrollment *(% of children)*	86.30	135	
Secondary school attainment *(% of population)*	42.80	119	
Gender parity in secondary attainment *(distance from parity)*	0.14	103	
Access to quality education *(0=unequal; 4=equal)*	0.88	147	
Access to Information & Communications	70.50	87	
Mobile telephone subscriptions *(subscriptions/100 people)*	146.92	1	
Access to online governance *(0=low; 1=high)*	0.68	75	
Media censorship *(0=frequent; 4=rare)*	3.33	35	
Internet users (% of pop)	33.82	135	
Health and Wellness	62.54	92	
Life expectancy at 60 *(years)*	21.58	66	
Premature deaths from non-communicable diseases *(deaths/100,000)*	280.70	52	
Access to essential services *(0=none; 100=full coverage)*	70.38	88	
Access to quality healthcare *(0=unequal; 4=equal)*	0.87	147	
Enviromental Quality	74.08	110	
Outdoor air pollution attributable deaths *(deaths/100,000)*	28.63	94	
Greenhouse gas emissions *(total CO2 equivalents)*	12.80	60	
Particulate matter	24.08	107	
Biome protection	7.02	140	

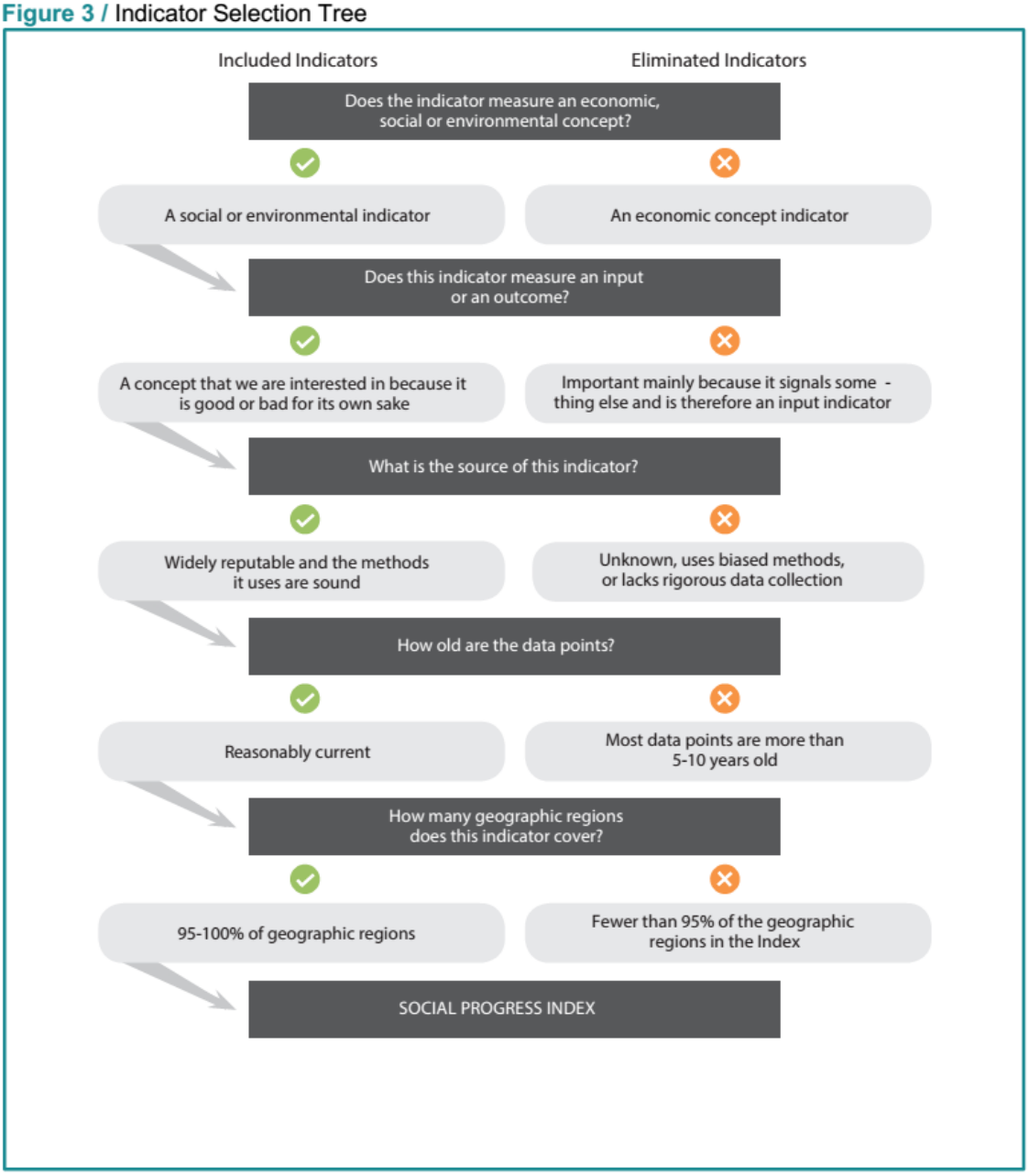
OPPORTUNITY	Score/Value	Rank	Strength/Weakness
	55.26	93	
Personal Rights	75.38	78	
Political rights *(0=no rights; 40=full rights)*	32.00	72	
Freedom of expression *(0=no freedom; 1=full freedom)*	0.82	64	
Freedom of religion *(0=no freedom; 4=full freedom)*	3.82	25	
Access to justice *(0=non-existent; 1=observed)*	0.24	156	
Property rights for women *(0=no right; 5=full rights)*	4.49	67	
Personal Freedom & Choice	63.12	95	
Vulnerable employment *(% of employees)*	34.19	90	
Early marriage *(% of women)*	15.61	138	
Satisfied demand for contraception *(% of women)*	80.90	41	
Corruption *(0=high; 100=low)*	34.00	112	
Inclusiveness	43.56	98	
Acceptance of gays and lesbians *(0=low; 100=high)*	0.45	37	
Discrimination and violence against minorities *(0=low; 10=high)*	5.80	90	
Equality of political power by gender *(0=unequal power; 4=equal power)*	1.92	87	
Equality of political power by socioeconomic position *(0=unequal power; 4=equal power)*	1.46	126	
Equality of political power by social group *(0=unequal power; 4=equal power)*	1.66	126	
Access to Advanced Education	38.99	121	
Expected years of tertiary education	1.61	85	
Women with advanced education *(%)*	0.31	140	
Quality weighted universities *(points)*	5.00	91	
Citable documents	0.02	173	

SOCIAL PROGRESS IMPERATIVE

Source**:** Social Progress Imperative, 2020b.

Appendix C: Social Progress Index indicator selection tree

Figure 3 / Indicator Selection Tree



Source: Stern et al., 2020, p. 8.