The International Civil Service Effectiveness (InCiSE) Index

InCiSE+

Technical Report

2019

InCiSE

Please contact incise@bsg.ox.ac.uk for any queries.

The International Civil Service Effectiveness (InCiSE) Index project is a collaboration between the Blavatnik School of Government and the Institute for Government. It has been supported by the UK Civil Service and funded by the Open Society Foundations.

Two reports have been published covering the 2019 edition of the InCiSE Index:

The Results Report provides an overview of the overall results of the InCiSE Index, and a summary for each country covered by the Index.

The Technical Report provides the methodology of the InCiSE Index. It includes details of the data sources, transformations, imputation, and weightings.

Contents

List of Tables

List of Figures

Reader's guide

In some tables and graphs countries are referred to by their ISO 3166-1: alpha-3 three-letter country codes. The codes for the 38 countries covered by InCiSE are:

AUS	Australia		**ISL**	Iceland
AUT	Austria		**ISR**	Israel
BEL	Belgium		**ITA**	Italy
BGR	Bulgaria		**JPN**	Japan
CAN	Canada		**KOR**	South Korea
CHE	Switzerland		**LTU**	Lithuania
CHL	Chile		**LVA**	Latvia
CZE	Czechia		**MEX**	Mexico
DEU	Germany		**NLD**	The Netherlands
DNK	Denmark		**NOR**	Norway
ESP	Spain		**NZL**	New Zealand
EST	Estonia		**POL**	Poland
FIN	Finland		**PRT**	Portugal
FRA	France		**ROU**	Romania
GBR	United Kingdom		**SVK**	Slovakia
GRC	Greece		**SVN**	Slovenia
HRV	Croatia		**SWE**	Sweden
HUN	Hungary		**TUR**	Turkey
IRL	Ireland		**USA**	United States of America

The following acronyms are used in some tables to refer to the 12 InCiSE indicators:

CAP	Capabilities		**INC**	Inclusiveness
CRM	Crisis and risk management		**INT**	Integrity
DIG	Digital services		**OPN**	Openness
FFM	Fiscal and financial management		**POL**	Policy making
			PRO	Procurement
HRM	Human resources (HR) management		**REG**	Regulation
			TAX	Tax administration

Chapter 1: Introduction

This report sets out the methodology for the 2019 edition of the International Civil Service Effectiveness (InCiSE) Index project. It provides an explanation of the high-level principles underlying the development of the InCiSE Index, the overarching methodology for the Index's calculation, the methodology for each of the constituent indicators that make up the Index, and documents the methodological changes made following the 2017 Pilot edition of the index.

1.1 Why InCiSE is needed

An effective civil service can play a vital role in determining a country's progress and prosperity. But what constitutes an "effective civil service"? The functions of the central government are not always directly comparable to other organisations in a given country. Thus, international comparisons of government and civil service activity are often sought.

InCiSE aims to define "effectiveness" more extensively than previous literature, drawing on a wide range of existing international data sources to bring together a set of indicators, each measuring a different dimension of civil service effectiveness. These indicators are then used to produce a composite (overall) score.

This creation of a new and concise set of civil service effectiveness indicators therefore serves as:

- An accountability tool: allowing citizens, government officials, and politicians to establish clearly and concisely how well their civil service is performing.

- A performance improvement tool: enabling senior decision makers to see the countries which perform best in each area, and therefore learn from them.

InCiSE has been developed following a literature review and in consultation with many experts, including academics from schools of government, think-tanks that monitor government effectiveness, international organisations, senior civil servants (past and present) and subject experts. InCiSE has also been the subject of an independent, international peer review process. The 2019 edition of InCiSE has also benefited from the feedback collected and provided since the publication of the 2017 Pilot.

1.2 Defining the civil service

Civil service effectiveness is well recognised in academic, international and practitioner communities as a highly complex area for analysis. As well as data limitations and the need to take account of country context factors, analysts are also faced with differing views on the definitions of both "civil service" and "effectiveness". The scope, responsibilities, and structure of the civil service vary across countries, creating the need to establish exactly what is being assessed, and how.

In defining the civil service there are a number of possible approaches to take:

■ First, a civil service can be defined by function: a narrow view of the civil service through this definition focuses on the central, "upstream" agencies which set policy direction and procedural regulation for "downstream" agencies. The broader view encompasses agencies responsible for service delivery.

■ Second, a civil service can be defined by national accounts: this perspective sees the civil service as made up of entities which are owned by the government, and whose financial reporting places them within the System of National Accounts (SNA) category of General Government.

■ Third, a civil service can be defined by employment regimes: under this definition, civil service entities are limited to those which are required to hire most employees under the civil service law, and those using other legal employment regimes are excluded.

However, conceptual and practical problems arise under each of these definitions. For example, staff commonly referred to as "civil servants" do not always have legally distinct employee contracts; the SNA definition is inconsistent with the views of many practitioners and researchers; and each alternative conception results in a large and unwieldy group of agencies.

InCiSE therefore takes a fourth and alternative approach, defining the scope of 'civil services' by outlining and measuring performance on the core functions of civil services; the parts which can generally be classified as civil service in every country. This approach leads to a focus on (i) functions which deliver services or affect citizens directly and (ii) public management and policy functions carried out in the centre of government.

The unit of analysis of interest for the InCiSE Index is the civil service, rather than the public sector more generally. InCiSE also focuses on civil service at the central/federal level – the highest level of government in a country/state – rather than at the regional or local level. Even with these parameters, isolating civil service performance with currently available data is still difficult, particularly given the varying sizes and shapes of civil services internationally.

Table 1.2.A gives more detail about what is included and excluded in the InCiSE Index.

Table 1.2.A   Scope of the InCiSE Framework

Part/function of the public sector	Degree of inclusion in the InCiSE framework
Civil service functions that deliver services to citizens and organisations directly (e.g. tax and social security administration at the central/ federal level).	A primary focus of the InCiSE framework
Public administration functions of central government (e.g. fiscal management, policy making, regulation)	A primary focus of the InCiSE framework
'Mission support' functions (e.g. HRM and procurement) that support the operation of central government organisations.	A primary focus of the InCiSE framework
Parts of the civil service which direct and support the wider public sector on specific policy areas (e.g. ministries of health or education) but may not deliver services to citizens directly.	Performance captured through the assessment of central government's public administration functions (e.g. policy making, regulation). Performance of policy areas themselves (e.g. quality of healthcare, educational attainment) are not assessed as these are not always the responsibility of central/federal government, moreover the policy goals and policy approaches taken are determined by political decision making.
Sub-national government/public administration (e.g. regional or local government)	While in some jurisdictions employees of sub-national governments may be classed as civil servants (e.g. via employment law) the scope of InCiSE is principally with the central/federal level of government in a country/state. However, general government/public administration (incorporating both central and sub-national government) may be used as a proxy where no central-level civil service data is available.
The wider public sector (e.g. schools, hospitals, police forces).	Out of scope. However, public sector data may be used as a proxy where no central-level civil service data is available.

1.3 The InCiSE framework

The purpose of the InCiSE framework is to define a common approach for assessing the effectiveness of a civil service, in a way which could realistically enable international data to be collected to measure against it. Whilst there are many alternative ways to define civil service effectiveness, the framework outlined here is informed by evidence and set out in such a way that if a civil service scores highly against it, it is reasonable to conclude that this civil service is high-performing relative to its international counterparts.

Our approach to deriving a common framework was to:

- Specify and adhere to a set of principles to inform the development of the framework:

  - Coherent – identifying the key elements and drivers of effective public administration

  - Comprehensive – covering all relevant aspects and drivers of the performance of public administration

  - Actionable – offering genuine insights into what drives excellent public administration that can be implemented

  - Transparent – a clear methodology and assessment process to ensure credibility, robustness, and replicability

  - Feasible – it is possible to collect data for a large group of countries at reasonable cost

- Draw on evidence to identify key features of a draft framework which was then extensively tested through consultation.

- Build on existing indicators and data where possible, striving to develop a more comprehensive framework capturing all aspects of civil service effectiveness.

- Refine the framework through consultation with a number of experts, including academics, think-tanks, international organisations, civil servants (both past and present) and subject experts.

A common approach for assessing organisational effectiveness is to think in terms of inputs, outputs, and outcomes. However, this is appears less attractive when considering civil services and the public administration-type functions they provide. While output and outcome measures may have the advantage of cutting through conceptual uncertainty, they can be problematic in this area for three reasons:

1. Outputs and outcomes can be affected by external factors, making it difficult to isolate the contribution of the civil service.

2. Measuring outputs and their value can be methodologically problematic, particularly as many public sector outputs are provided free at the point of consumption.

3. Focusing on outputs and outcomes means that normative and procedural concerns which are also relevant to effectiveness can be ignored.

Given these concerns, the preferred approach here is to focus on the effectiveness of the procedures within the civil service which (often indirectly) affect outcomes. The framework's approach is therefore more process focused and output focused, as outlined in Figure 1.1. An advantage of choosing process-based indicators is that they are more instructive

for potential performance improvements – it is processes that are ultimately changed to increase effectiveness.
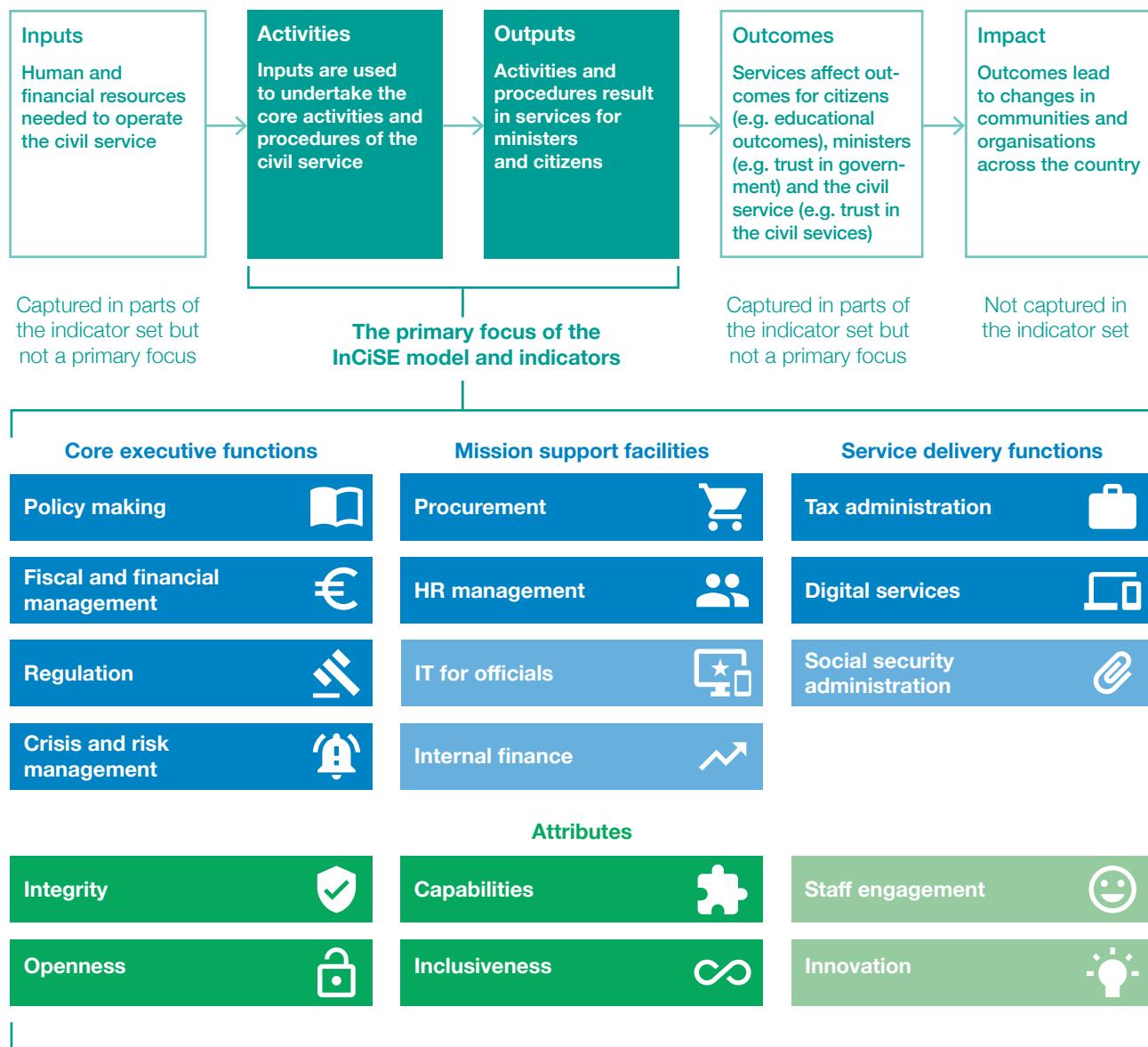
Although procedural definitions also come with problems (they may not actually correlate with positive outcomes, for example) certain procedural measures remain at the core of any measure of effectiveness. Where there is evidence to support the relationship between procedures and positive outcomes, procedures may also be intrinsically beneficial. For example, meritocracy of recruitment procedures in the civil service are important because there is broad agreement that such procedures and outcomes are associated with an effective civil service. However, the extent to which recruitment processes reward merit is also important in the principle of fairness which is valued in itself.

The InCiSE framework, shown in Figure 1.1, defines the core characteristics of an effective civil service. To do this, it assesses effectiveness on the basis of two interrelated dimensions: 1) the delivery of its core functions and 2) an underlying set of attributes which are important drivers of effectiveness across all parts of the civil service. Collectively the functions and attributes are called 'indicators' within the InCiSE model. Section 1.6 describes in more detail how the framework is implemented as a statistical model.

Functions: On one side, civil services deliver a set of central executive functions for ministers. These may help to formulate policy for the country (the effects of which are borne by citizens). On the other side, the services interact more directly with citizens through the delivery of services such as tax administration. Finally, in the centre, supporting these core external functions, are mission support functions such as HR management or IT services for officials. By looking across all three types of function, the aim is to measure how well civil services deliver the core elements of their roles. The functions identified by the InCiSE model are:

- Policy making: The quality of the policy making process, including how policy is developed and coordinated across government and monitored during implementation.

- Fiscal and financial management: The quality of the budgeting process and the extent to which spending decisions are informed through economic appraisal and evaluation.

- Regulation: The extent and quality of regulatory impact assessments and the degree of stakeholder engagement involved in them.

- Crisis and risk management: The effectiveness with which the government engages the whole of society to better assess, prevent, respond to and recover from the effects of extreme events.

- Procurement: The extent to which the procurement process is efficient, competitive, fair, and pursues value for money.

Figure 1.1   The InCiSE Index Framework

Inputs	Activities	Outputs	Outcomes	Impact
Human and financial resources needed to operate the civil service	Inputs are used to undertake the core activities and procedures of the civil service	Activities and procedures result in services for ministers and citizens	Services affect outcomes for citizens (e.g. educational outcomes), ministers (e.g. trust in government) and the civil service (e.g. trust in the civil sevices)	Outcomes lead to changes in communities and organisations across the country

Captured in parts of the indicator set but not a primary focus

The primary focus of the InCiSE model and indicators

Captured in parts of the indicator set but not a primary focus

Not captured in the indicator set

Core executive functions

Policy making	📖
Fiscal and financial management	€
Regulation	
Crisis and risk management	🔔

Mission support facilities

Procurement	🛒
HR management	👥
IT for officials	
Internal finance	📈

Service delivery functions

Tax administration	💼
Digital services	
Social security administration	📎

Attributes

| Integrity | ✅ |
| Openness | 🔒 |

| Capabilities | 🧩 |
| Inclusiveness | ∞ |

| Staff engagement | 🙂 |
| Innovation | 💡 |

IT for officials, internal finance, social security administration, staff engagement, and innovation are not currently measured in the InCiSE Index.

- HR management: The meritocracy of recruitment and the extent to which civil servants are effectively attracted, managed and developed.

- IT for officials: The extent to which civil servants have the technology and digital tools to work efficiently.

- Internal finance: The extent to which civil service operations are supported by well-managed and efficient finance systems, particularly on the alignment of finance with the business strategy and the level of civil servant satisfaction with finance support.

- Tax administration: The efficiency and effectiveness of tax collection (at the central/ federal level).

- Social security administration: The efficiency and effectiveness of social security administration (at the central/ federal level).

- Digital services: The availability and usability of national-level digital public services.

Attributes: Every civil service also has an underlying set of attributes which are important drivers of how effectively they deliver core functions. These attributes should apply to all parts of the civil service and are not specific to particular parts or functions. The inclusion of attributes in the framework is based on both a normative and a positive judgement: civil services should aim to cultivate and demonstrate these attributes as they are commonly (but not necessarily universally) understood as aspects of best practice, and the included attributes should generally be determinants of performance across all functions.

- Integrity: The extent to which civil servants behave with integrity, make decisions impartially and fairly, and strive to serve both citizens and ministers.

- Openness: The regular practice and degree of consultation with citizens to help guide the decisions we make and extent of transparency in our decision-making.

- Capabilities: The extent to which the workforce has the right mix of skills.

- Inclusiveness: The extent to which the civil service is representative of the citizens it serves.

- Staff engagement: Staff levels of pride, attachment and motivation to work for their organisation.

- Innovation: The degree to which new ideas, policies, and ways of operating are able to freely develop.

The 2019 edition of InCiSE measures 12 of the 17 functions and attributes defined by the framework. Chapter 3 provides further detail of the definition and measurement of each of these indicators. Four of the five indicators (IT for officials, internal finance, staff engagement, and innovation) are not included because it has not been possible to identify suitable or sufficient data for cross-country measurement. One of the five indicators (social security administration) was measured in the 2017 Pilot but has been depreciated due to data quality concerns.

1.4 The InCiSE data model

The InCiSE Index is based on a framework that describes the various components of an effective civil service. The Index operationalises this framework by measuring a series of indicators that correspond to the different components of the InCiSE framework. The overall InCiSE Index results are a composite of the indicator scores. In turn the indicators are split into themes, which describe important sub-divisions of the indicator. Scores for these themes are not computed but the theme structure is part of the weighting used in the calculation of the indicator scores. The themes within an indicator are represented by individual metrics, which ideally measure tangible qualities of the civil service that can be acted upon or influenced by senior officials. Most of the InCiSE metrics are single data points published by the data source providers, however some metrics are calculated from multiple data points. Figure 1.2 outlines the "data model" used by InCiSE, showing how individual data points from the external data sources combine to form the metrics, indicators and composite index of InCiSE.

InCiSE is not intended to measure inputs (e.g. money/resources) or public policy outputs (e.g. unemployment benefits paid; taxes collected) or citizen outcomes (e.g. life expectancy, GDP per capita, citizen wellbeing), as these are typically determined by political decisions about the size of the state and what it is aiming to achieve. Rather, InCiSE is designed to assess the effectiveness of the way in which the civil service of a country uses the inputs it has been given to deliver the policy outputs/ outcomes that it has been set.

One of the main aims of the Index is to provide a mechanism for civil services to learn from each other: in particular to offer a data-driven approach to identify sources of good practice. To achieve this, InCiSE does not assess the absolute performance of different civil services. Instead, it converts the absolute performance captured in the individual metrics into relative assessments of performance of the countries included in the Index. This means that scoring poorly in InCiSE does not in itself indicate absolute poor performance, rather that when compared to other countries performance is lower. Similarly, scoring well in InCiSE does not in itself indicate absolute high performance, but that when compared to other countries performance is higher. The fact that there is no natural scale for civil service performance strengthens the case for measuring relative rather than absolute performance.
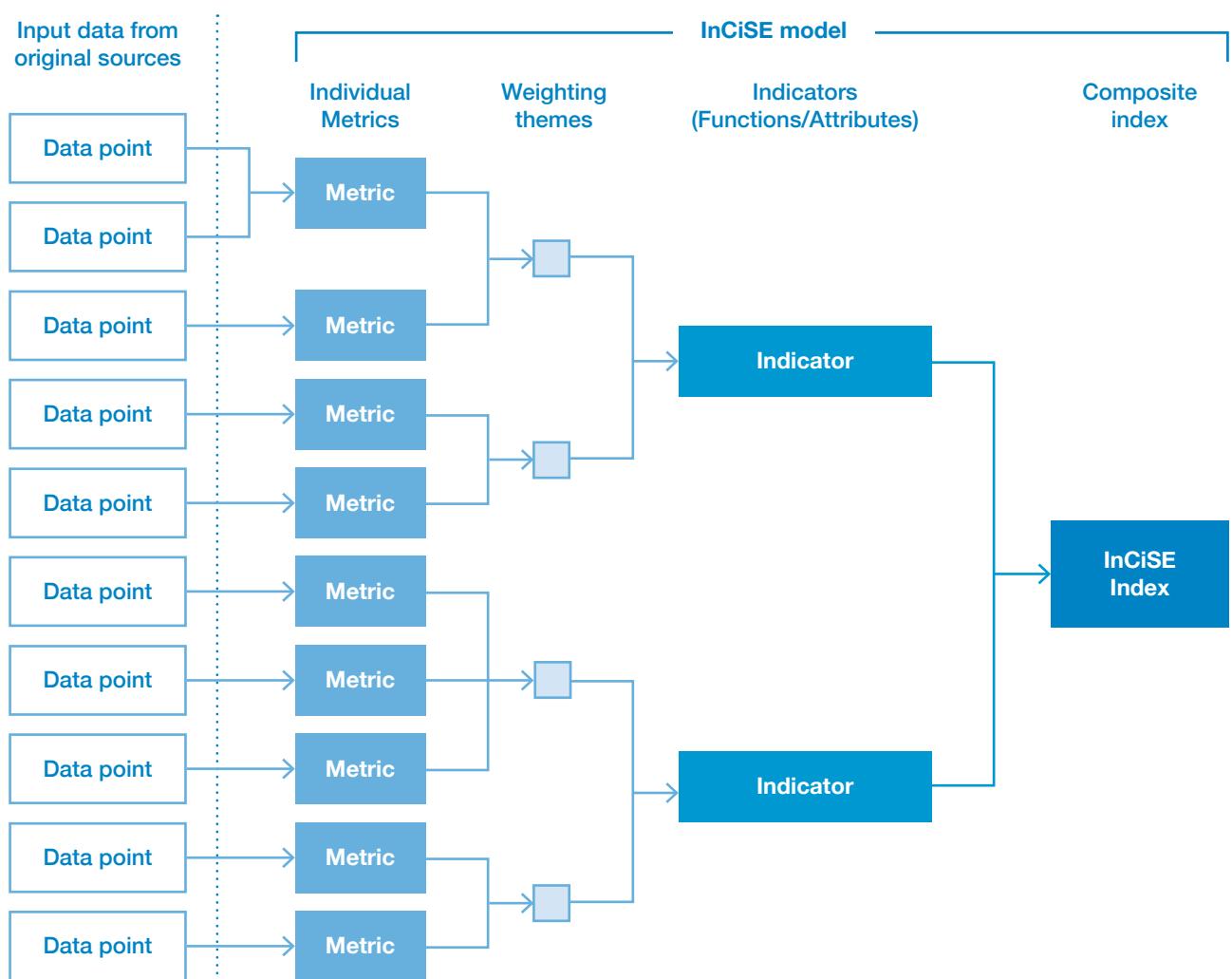
1.5 Eligibility of metrics

Metrics are eligible for inclusion in InCiSE if they meet the following criteria:

- **The data must be published in a free-to-access form** in the public domain and online. That is, an independent person must be able to access the data from a publicly accessible and free-to-use website.

- **The data must be actionable.** That is, the data must measure some quality or component of the civil service that government officials and ministers can act on to improve performance. Where data for the civil service is not available, public sector proxies can be used, but these must still be data that represent something that can be acted on.

- **The data must be quantifiable**, and if not directly collected and published as numerical data there must be a way to convert the data into a clear and relevant numerical format.

Figure 1.2   The InCiSE data model

1.6 Technical approach

The technical approach for the 2019 model has used the 2017 InCiSE pilot edition of the model as the reference point for its methodology, however the data collection and statistical model was rebuilt from first principles to provide a "clean slate" for the 2019 modelling. That is, the 2019 model did not start as a copy of the 2017 final model with data updated to reflect the latest values with new data inserted and code amended. Instead, the 2019 model has been developed from scratch using R (rather than the mix

of Excel and Stata used for the 2017 Pilot). This approach has been taken to (i) minimise the potential of error and improve quality assurance processes, and (ii) improve the openness, reproducibility and extensibility of the InCiSE model. The approach adopted for the 2019 InCiSE model is based on the *Reproducible Analytical Pipelines* approach developed by data scientists at the UK Government Digital Service (Gregory and Upson, 2018). The technical approach to coding and data management/processing was also influenced by the *tidyverse* principles

(Wickham and Grolemund, 2017; Wickham, 2015). A full list of the software packages used to develop and implement the modelling are listed in the References at the end of this report.

1.7 Quality considerations and limitations

As with any analytical endeavour, there are limitations to how far and in what ways the InCiSE Index can and can't be used. Furthermore, given its early stage of development, the InCiSE Partners are clear that the index remains an experimental methodology that is subject to change and evolve in order to refine and improve the Index.

This section outlines some of the key considerations that should be taken into account when reviewing and using the InCiSE Index. Stating these limitations is not to downplay the value of the index as a tool for cross-country comparison, rather it is to help users understand the data they are using. Furthermore, InCiSE is not intended to be used in isolation but to enhance the range of evidence available about government effectiveness. Users should build a "rich picture" of the situation by triangulation across the results from InCiSE, the underlying results from InCiSE's source metrics or other international comparisons, and domestic information for which there is no international comparisons.

There are a number of different aspects that should be taken into consideration in regards to the quality of data used in InCiSE:

- Recency and frequency of the data: InCiSE 2019 uses the most recently available data as at 30 November 2018. Some metrics in InCiSE are collected annually, others biennial or longer, or are ad-hoc in their repetition. As a result, some metrics may use data that does

not accurately reflect the most recent situation.

- Depth of the data: Some metrics represent a single measure in a survey, some are aggregations of multiple measures by the InCiSE model, while some are composite indicators compiled by others parties that are based on a range of metrics.

- "Spill over": Some measures that contribute to one of the InCiSE indicators may be relevant to other indicators, but wherever possible this has been avoided. No original piece of data used by the InCiSE model is used more than once in order to ensure that the overall figures are not overly-reliant on a particular data source.

- Public sector proxy: The purpose of the InCiSE Index is to measure the effectiveness of a country's national and central civil service. However, some metrics measure the performance of the public sector at large – or at least a larger subset than the specific unit of analysis that InCiSE is interested in. In this case the public sector measures can only be considered proxies.

- Proxy measures of effectiveness: The true nature of the effectiveness of a country's civil service is inherently unobservable, and cannot be comprehensively observed in an empirical study. The purpose of the InCiSE project is to provide a means to combine a range of proxy measures to provide insight into the effectiveness of civil services. The InCiSE framework enables this analysis by providing a way to conceptualise how a civil service operates. The model therefore uses measures about the functions and attributes of a civil service to produce an estimate of effectiveness.

1.8 Relationship with other indicators and data collections

In setting the civil service as our unit of interest, it is also important to distinguish the difference of InCiSE with other 'governance' indicators (particularly the World Bank's Worldwide Governance Indicators and the Bertelsmann Foundation's Sustainable Governance Indicators). Other governance indicators take a broad view of the topic of governance, including assessments of political decision making within governing parties, the quality of democracy, the ability to hold the government to account, and the freedoms of media and civil society. These are important factors in considering the governance of a country in general. InCiSE seeks to complement these 'broad' assessments of governance by providing a deeper investigation with a narrower focus on a key element of the operation of government – the civil service.

Besides 'broad' governance indicators there are also thematic indicators that focus on specific elements of governance; for example, the World Wide Web Foundation's Open Data Barometer, the World Justice Project's Rule of Law Index or the OECD's regulation indicators. There are also indicators focused on other themes that cut across sectoral boundaries (for example Transparency International's Global Corruption Barometer, or the World Economic Forum's Doing Business Report) which contain a large amount of information about countries but where only a few measures directly relate to central government/civil service performance.

Finally, there are also a range of data collections made by international organisations and other institutions (notably the OECD, the European Commission, and the United Nations) about the functioning of government/the civil service but which do not produce single composite assessments.

The InCiSE framework and index has been designed and developed to re-use data from these indicators and data sources to produce a single coherent and comparable data model that allows a wide variety of parties interested in civil service reform to make a high-level assessment of how the civil services of different countries compare. The InCiSE Index should not be used in isolation, but in combination with reference to the source datasets as well as with domestic data from within a country about performance across the various indicators.

Further considerations about the specific data quality of the InCiSE data and results is provided in Chapter 2.

1.9 Structure of this report

This Technical Report on the InCiSE Index is intended to describe the methodology, data and limitations of the approach used. The results of the Index can be found in the accompanying 2019 Main Report. Including the introductory chapter, there are seven chapters in this report:

- Chapter 2: Methodology of the InCiSE Index outlines the data processing, calculation of the InCiSE indicators, and calculation of the InCiSE Index.

- Chapter 3: Methodology of the InCiSE indicators sets out the methodology for each of the 12 indicators that make up the 2019 index.

- Chapter 4: Summary of changes from the 2017 Pilot highlights the changes made in within the methodology of each indicator, as well as in the overarching methodology of the index.

- Chapter 5: Sensitivity analysis describes some of the uncertainties associated with the modelling process and subjective choices, and the consequent impact on the Index results.

- Chapter 6: Future development sets out the next steps for future consideration and development of the index methodology.

There are also two annexes to the report that provide additional detail:

- Annex A: Composite metrics provides details of how the different composite metrics used in the InCiSE Index have been constructed.

- Annex B: Sensitivity analysis results provides detailed results of the different tests conducted as part of the sensitivity analysis

Chapter 2: Methodology of the InCiSE Index

As outlined in Chapter 1, the InCiSE Index is a composite index formed from a series of indicators, each of which is comprised of a set individual metrics. The overall Index is the normalised and weighted average of the scores of the constituent InCiSE indicators. The InCiSE indicators are themselves normalised weighted averages of their individual metrics. The calculation and modelling process to produce the Index is as follows:

1. Data processing:
   a. Data preparation [section 2.1]
   b. Data quality assessment [2.2]
   c. Country coverage selection [2.3]
   d. Imputation of missing data [2.4]
   e. Data normalisation [2.5]

2. Calculation of the InCiSE indicators [2.6]:
   a. Raw score calculated as a weighted average of the individual metrics
   b. Raw score normalised to produce final indicator score

3. Calculation of the InCiSE Index [2.7]:
   a. Raw score calculated as a weighted average of the indicator scores
   b. Raw score normalised to produce final Index score

This chapter outlines the methodology for each of these different stages, while Chapter 3 provides details on the specific methodology of each of the InCiSE indicators.

2.1 Data preparation

The data for InCiSE comes from a wide range of independent sources, such as the UN's E-Government Survey, Transparency International's Global Corruption Barometer, and Bertelsmann's Sustainable Governance Indicators (SGIs).[1] The InCiSE partnership does not produce any of the source data itself or engage in primary data collection. The data for the 2019 edition of InCiSE is the latest available as of 30 November 2018. As well as the source metrics some additional data are collected to aid in the imputation of missing data – this data does not directly contribute to the scores and therefore is not included in the published results.

Some of the source data requires processing before it is suitable for use in the InCiSE calculations and modelling. For example:

- Binary/multiple categorical data: some of the source data are binary measures (e.g. yes/no questions) or assess multiple categories (e.g. groups subject to whistleblower protection). In many cases this data is summed.

- Individual level microdata: InCiSE uses a custom analysis of the Programme for the International Assessment of Adult

---

1   A full list of data sources can be found in the References chapter at the end of this report.

Competencies (PIAAC) individual-level microdata to produce country scores. The Opentender data on procurement is on individual contracts, which also requires analysis to produce country scores.

- Negatively framed data: Some of the source data is based on negatively framed questions, where a higher score is poorer performance than a lower score. To align with other metrics, this data is inverted so that higher scores relate to better performance than lower scores.

- Calculations against reference data: For the inclusiveness indicator, women's representation in the civil service/public sector is compared to the labour market in general. Tax administration from the OECD is published as raw data. InCiSE uses rates based on these data which must therefore be calculated.

Chapter 3 outlines the underlying source data for each of the indicators, and covers the specific transformations that are applied to the source data. Annex A outlines the construction and calculation of composite metrics.

When importing data to the InCiSE model, data is matched against a reference list of 249 countries and territories produced by Arel-Bundock et al (2018) using the 3-digit ISO 3166-1 alphanumeric codes. Some source data natively uses the 3-digit ISO country codes, but some use the 2-digit ISO code, another code system, or a name of the territory (either the official long/short name, or colloquial name). Therefore, as part of data preparation, all country references are converted to the 3-digit ISO country code.

2.2 Data quality assessment

In order to provide a clearer understanding of the quality of the InCiSE Index, a data quality assessment has been calculated and published alongside the 2019 edition. This assessment has a dual role: it is an important piece of metadata that will help users of the InCiSE Index better understand the results, but it has also been used to determine the country coverage of the InCiSE Index. This section describes the method for conducting the data quality assessment. The use of the assessment for country selection and weighting are discussed in sections 2.3 and 2.7 respectively, while a wider discussion of data quality based on the results of the assessment is provided at section 2.8.

The data quality assessment is a purely quantitative exercise based on three factors: data availability, the (non-)use of public sector proxy data, and the recency of the data. The assessment does not include any subjective evaluation of the methodology or the quality of the data sources that the underlying data used by InCiSE comes from. The data quality assessment also does not incorporate assessments of the reliability or validity of indicator and index construction. Its purpose is to provide an assessment of easily quantifiable characteristics of the data, which can help interpretation of the InCiSE results for countries and of the indicators.

The simple mean of the three measures is taken as the data quality score for each country for each indicator. The 12 overall indicator quality scores are then combined as a simple mean score to produce an overall data quality assessment for each country.

The data quality assessment is calculated for each indicator within each country, then averaged to produce an overall score for each country. For each indicator, the data quality assessment is based on three

measures: (1) the proportion of metrics with data; (2) the proportion of metrics that have civil service specific data; and (3) the recency of the data. All three measures take a simple assessment of whether data is missing or present as their basis. However, each measure has different weighting rules for the data:

- Data availability: A missing data point for a metric with a within-indicator weight of 15% will give a greater penalty than a missing data point for a metric with a within-indicator weight of 5%.

- Civil service data (1) or a public sector proxy (0): Data points that come from public sector data are treated as equivalent to being missing.

- Recency of the data: The reference year of the metric is scaled from 0 (for 2012 the earliest year) to 1 (for 2018 the latest year) and used as the weighting.[2]

The country indicator data quality scores and overall data quality assessment ($DQA_{c,i}$) for a given country (c) and indicator (i) is calculated by multiplying the missing data matrix of the metrics in the indicator for that country ($d_{c,i}$) by each of: the within indicator weighting for the metrics in the indicator (m_i), the proxy data status of each metric in the indicator (s_i), the recency of each metric in the indicator (r_i). The resulting products are summed and divided by three to give the mean data quality for that country and indicator.

$$DQA_{c,i} = \frac{(d_{c,i} \cdot m_i) + (d_{c,i} \cdot s_i) + (d_{c,i} \cdot r_i)}{3} \quad \dots (1)$$

The overall data quality indicator for a country (DQA_c) is then calculated as the sum of data quality assessment scores of that country for each indicator ($\sum DQA_{c,i}$) divided by the number of indicators (n_i)

$$DQA_c = \frac{\sum DQA_{c,i}}{n_i} \quad \dots (2)$$

The data quality assessment scores therefore have a theoretical range from 0 to 1. 0 represents there being no metrics available and 1 represents there being data for all metrics, with all data representing the civil service (i.e. not providing a public-sector proxy) and all data relating to the latest available year. Table 2.2.A illustrates the complex picture of data quality across all countries and indicators.

The table shows how maximum data quality varies from 0.333 for capabilities, where the available data is for a public sector proxy and the oldest data in the model, to 1.000 for policy making, where all the available data relates to the civil service and is at the latest available data.

The indicators for openness, fiscal & financial management and crisis & risk management have good data quality (DQA score greater than or equal to 0.5) for a very large number of countries. Other indicators (such as HR management or tax administration) have a moderate number of countries with good data quality, but have a large number of countries with poorer data quality. Finally, some indicators (such as digital services or policy making) have data for only a small number of countries, which is typically due to the source data covering only OECD or EU members (or both).

---

2   For example a datapoint with a reference year of 2013 will be weighted 0.1667, while one with a reference year of 2016 will be weighted 0.6667

Table 2.2.A   Data quality assessment (DQA) results across the 12 InCiSE indicators and overall, for all 249 countries and territories considered by the InCiSE data model

Indicator	Highest country DQA score	Country distribution of DQA scores		
		DQA ≥ 0.5	0.5 > DQA > 0	DQA = 0
Capabilities	0.333	0	31	218
Crisis & risk management	0.855	95	13	141
Digital services	0.581	34	0	215
Fiscal & financial management	0.889	109	88	52
HR management	0.673	37	83	129
Inclusiveness	0.722	34	82	133
Integrity	0.569	30	127	92
Openness	0.928	105	93	51
Policy making	1	41	0	208
Procurement	0.722	20	24	205
Regulation	0.963	38	5	206
Tax administration	0.852	46	141	62
Overall data quality assessment	**0.757**	**38**	**162**	**49**

2.3 Country coverage selection

For the 2017 Pilot edition of the InCiSE Index only two countries had data for all 76 metrics, and a simple threshold of 75% data availability plus membership of the OECD were used as the selection criteria for country availability. However, analysis of the pilot showed (as Table 2.2.A shows) that there is a mixed picture of data availability and quality across indicators which is not reflected in this simple threshold. The data quality assessment outlined in section 2.2 provides a more nuanced way to consider the variation of data availability and quality, and is therefore used to determine which countries are included in the final version of the index for the InCiSE 2019.

In determining country coverage, the InCiSE Partners have decided to use an overall data quality assessment score of 0.5 or greater for the threshold for country inclusion. 38 countries reached this score. Although two further countries would be included if data quality scores were rounded to 1 decimal place, these two countries have lower data availability (57% and 51% of all metrics respectively), which is judged to be too low for reliable analysis. Therefore, the 38 countries with a data quality score of 0.5 or higher (when rounded to 2-decimal places) are included in the 2019 edition of the InCiSE Index. This includes all 31 countries covered by the InCiSE pilot.

Table 2.3.A Data quality assessment (DQA) results for the 38 countries included in the 2019 index

Indicator	Lowest country DQA score	Highest country DQA score	Mean country DQA score	Country distribution of DQA scores		
				DQA ≥ 0.5	0.5 > DQA > 0	DQA = 0
Capabilities	0	0.333	0.244	0	38	10
Crisis & risk management	0	0.855	0.631	26	12	1
Digital services	0	0.581	0.444	29	9	9
Fiscal & financial management	0.439	0.889	0.783	37	1	0
HR management	0.293	0.673	0.64	35	3	0
Inclusiveness	0.375	0.722	0.663	33	5	0
Integrity	0.402	0.569	0.526	29	9	0
Openness	0.283	0.928	0.818	35	3	0
Policy making	1	1	1	38	0	0
Procurement	0	0.722	0.513	20	18	2
Regulation	0.339	0.963	0.908	35	3	0
Tax administration	0.352	0.852	0.77	34	4	0
Overall data quality assessment	0.501	0.757	0.662	38	0	0
% of metrics available (2017 Pilot approach)	65%	100%	86%			

Table 2.3.A provides an overview of the country-level data quality scores for the group of 38 countries. The table shows that for most indicators the 38 countries have generally good data quality. However, for four indicators (capabilities, crisis & risk management, digital services and procurement) there are a small number of countries with no available data at all.

Table 2.3.B provides a summary of the data quality assessment for all 38 countries selected for the 2019 edition of InCiSE, plus the five countries with the next highest data quality score. One country (the United Kingdom) achieved the highest overall data quality score of 0.757, followed closely by five others (Italy, Poland, Sweden, Norway and Slovenia).

Table 2.3.B   Data quality assessment (DQA) results by country

Country		Overall DQA score	Percent of all metrics available	Number of indicators where: 0.5 > DQA > 0	Indicators with completely missing data (DQA = 0)	
GBR	United Kingdom	0.757	100%	1	0	
ITA	Italy	0.755	99%	1	0	
POL	Poland	0.755	99%	1	0	
SWE	Sweden	0.755	99%	1	0	
NOR	Norway	0.752	99%	1	0	
SVN	Slovenia	0.75	99%	1	0	
AUT	Austria	0.738	98%	1	0	
FIN	Finland	0.736	97%	2	0	
ESP	Spain	0.733	97%	1	0	
NLD	The Netherlands	0.731	98%	1	0	
FRA	France	0.718	97%	2	0	
PRT	Portugal	0.716	85%	1	1	CAP
DNK	Denmark	0.707	93%	2	0	
DEU	Germany	0.701	96%	2	0	
GRC	Greece	0.696	94%	2	0	
SVK	Slovakia	0.692	93%	1	0	
HUN	Hungary	0.671	81%	1	1	CAP
EST	Estonia	0.669	90%	2	0	
CZE	Czechia	0.659	91%	3	0	
TUR	Turkey	0.65	90%	4	0	
MEX	Mexico	0.648	73%	3	2	CAP, DIG
NZL	New Zealand	0.644	83%	4	1	DIG
CHL	Chile	0.643	79%	4	1	DIG
CAN	Canada	0.638	78%	4	1	DIG
KOR	Republic of Korea	0.636	78%	4	1	DIG

Further discussion on data quality issues are provided at the end of this chapter in section 2.8, covering both the quality of the indicators and interpretation of country level results from the InCiSE Index.

2.4 Imputation of missing data

As seen in Table 2.3.B only one country has complete data (i.e. 100% of metrics). The average level of data availability is 86% across the 38 countries, and 7 of the included countries have data availability below the 75% threshold used for the 2017 Pilot, with the lowest level of data availability being 65%. Of the 38 countries, 15 have one

Table 2.3.B (continued)

Country		Overall DQA score	Percent of all metrics available	Number of indicators where: 0.5 > DQA > 0	Indicators with completely missing data (DQA = 0)	
BEL	Belgium	0.635	85%	3	1	CRM
LVA	Latvia *[new]*	0.628	75%	2	1	CAP
CHE	Switzerland	0.627	79%	2	1	CAP
AUS	Australia	0.618	71%	3	3	CAP, DIG, PRO
LTU	Lithuania *[new]*	0.615	82%	5	0	
IRL	Ireland	0.614	84%	4	0	
JPN	Japan	0.597	75%	5	1	DIG
USA	United States of America	0.579	74%	4	2	DIG, PRO
ISR	Israel *[new]*	0.578	72%	5	1	DIG
ISL	Iceland *[new]*	0.563	68%	5	1	CAP
ROU	Romania *[new]*	0.529	66%	5	1	CAP
BGR	Bulgaria *[new]*	0.511	66%	6	1	CAP
HRV	Croatia *[new]*	0.501	65%	6	1	CAP
Mean of 38 countries		**0.635**	**82%**	**3.3**	**0.8**	
Countries with the next five highest data quality scores:						
COL	Columbia	0.471	57%	6	3	CAP, DIG, POL
LUX	Luxembourg	0.46	51%	7	2	CAP, INC
CYP	Cyprus	0.435	64%	9	1	CRM
CRI	Costa Rica	0.417	48%	7	3	CAP, DIG, POL
MLT	Malta	0.375	49%	9	2	CAP, CRM

[new] indicates countries included in the 2019 edition of the InCiSE Index that were not part of the 2017 Pilot.

indicator with a data quality score of 0 (i.e. no data at all for that indicator), two countries have two indicators with a data quality score of 0 and one country has three indicators with a data quality score of 0.

This presents issues for the analysis of the data and providing an effective method for aggregating the metrics into indicators and an overall index. The 2017 Pilot edition of InCiSE adopted two methods for imputation: multiple imputation using linear regression and median imputation. For the 2019 edition of InCiSE a decision has been made to move fully to a multiple imputation approach, using the 'predictive mean matching' (PMM) technique of van Buren's (2018) Multiple Imputation using Chained Equations (MICE) R software package. The PMM technique uses correlation – of both the values and pattern of missing data – to identify for a country with missing data those countries in the dataset that closely match it, and randomly select one of those to replace the missing value. Following the approach set out by van Buuren (2018), for each missing value 15 imputations are generated (each of which has also been iterated 15 times). A simple mean of these 15 imputation values is then calculated and used as the country's value in the 'final' dataset.

Imputation is handled on a per-indicator basis – in most cases imputation will be solely from within the metrics of that indicator. However, a few indicators have external predictors, either data from elsewhere in the InCiSE model or from an external data source. Full details of the imputation approach for each indicator is described in Chapter 3.

2.5 Data normalisation
As a result of coming from different sources, the underlying data that drives the InCiSE model has a variety of formats: some are proportions or scores from 0 to 1 or 0 to 100;

some are ratings on a scale, or the average of ratings given by a set of assessors/survey participants; and some are counts. The different formats of these data are not easily comparable, and cannot be directly averaged together to produce a combined score. In order to facilitate the comparison and combination of data from different sources, the metrics are normalised so that they are all in a common format.

There are a number of normalisation techniques that could be used. A useful discussion of the different methods is provided in the OECD (2008) Handbook on Constructing Composite Indicators. The InCiSE Index uses min-max normalisation at all stages, as this maintains the underlying distribution of each metric while providing a common scale of 0 to 1. The common scale is of particular benefit, as it helps achieve InCiSE's goal of assessing relative performance. In the min-max normalisation 0 represents the lowest achieved score and 1 represents the highest achieved score. It is therefore important to note that scoring 0 on a particular metric, indicator or the index itself does not represent poor performance in absolute terms, nor does scoring 1 represent high performance in absolute terms. Rather the country is either the lowest or highest performing of the 38 countries selected.

The min-max normalisation operates via the following mathematical formula:

$$m_c = \frac{x_c - x_{min}}{x_{max} - x_{min}} \ \dots \ (3)$$

For a metric for a given country its normalised score (m_c) is calculated as the difference of the country's original score (x_c) from the metric's minimum score (x_{min}) divided by the range of the metric's scores (the difference of the metric's maximum score (x_{max}) from the metric's minimum score (x_{min}).

2.6 Calculation of the InCiSE indicators

Once the data has been processed, missing data imputed, and the metrics normalised, the InCiSE indicators can be calculated. There are two stages to the calculation of the indicators: the weighting of the metrics into an aggregate score, and the normalisation of that score.

As outlined in Figure 1.2, the InCiSE data model first groups metrics into themes before aggregating into the indicator scores themselves. These themes are purely structural and scores for them are not computed. The raw score for an indicator follows this formula:

$$i_c = \sum \left(m_{i,c} \cdot w_m \cdot w_t \right) \dots (4)$$

A country's raw score for an indicator (i_c) is calculated as the sum of the product of each metric within the indicator for that country ($m_{i,c}$) with the weight of that metric within its theme (w_m) and the weight of that theme within the indicator (w_t). The weighting structure for each indicator is listed in detail in Chapter 3. After the raw scores are calculated they are normalised as described in section 2.5 above.

2.7 Calculation of the InCiSE Index

The InCiSE Index is an aggregation of the InCiSE indicators. Ideally, the indicators would be combined equally, however in producing the 2017 Pilot edition the InCiSE Partners felt it important to consider relative data quality. In the 2017 Pilot this was done by placing a lower weight on the indicators measuring 'attributes' than those measuring 'functions', as the four attribute indicators were considered to generally have lower data quality than those measuring functions. The 2019 edition builds on this approach to weighting by using the results of the data quality assessment (section 2.2).

For this approach to weighting, two-thirds of the weighting is allocated on an equal basis, while one third is allocated according to the outcome of the data quality assessment. The weight for an indicator is calculated as follows:

$$w_i = \left(\frac{2}{3} \cdot \frac{1}{n_i} \right) + \left(\frac{1}{3} \cdot Q_i \right) \dots (5)$$

Here the indicator weight (w_i) is equal to the product of two-thirds and the equal share (1 divided by n_i, the number of indicators; i.e. 1/12) plus the product of one-third and the data quality weight (Q_i). The data quality weight is calculated first by summing the data quality scores of the 38 selected countries for the indicator. The indicator's data quality sum is then divided by the sum of all indicator data quality scores, in essence providing a score that represents that indicator's share of the total data quality for the 38 countries selected. The resulting weights are shown in Table 2.7.A.

A country's overall raw index score (I_c) is thus calculated as the sum of the product of the normalised indicator scores for the country (i_c) with the indicator weights (w_i):

$$I_c = \sum i_c \cdot w_i \dots (6)$$

After calculating the raw index scores, they are then are normalised as outlined in section 2.5, resulting in the overall index scores for the 2019 edition of InCiSE.

Table 2.7.A   InCiSE indicator weightings

InCiSE indicator	Sum of data quality scores	Share of total data quality scores	Final weight (2/3 equal, 1/3 adjusted)	Approximate fraction
Capabilities	9.271	3.1%	6.6%	1/15
Crisis & risk management	23.967	7.9%	8.2%	1/12
Digital services	16.855	5.6%	7.4%	1/13
Fiscal and financial management	29.763	9.9%	8.8%	1/11
HR management	24.332	8.1%	8.2%	1/12
Inclusiveness	25.188	8.3%	8.3%	1/12
Integrity	19.995	6.6%	7.8%	1/13
Openness	31.100	10.3%	9.0%	1/11
Policy making	38.000	12.6%	9.8%	1/10
Procurement	19.500	6.5%	7.7%	1/13
Regulation	34.510	11.4%	9.4%	1/11
Tax administration	29.269	9.7%	8.8%	1/11
Overall	301.749	100.0%	100.0%	

2.8 Data quality considerations

Sections 2.3 and 2.7 illustrate how the data quality assessment described in section 2.2 are used within the InCiSE model for country selection and indicator weighting. The assessment can also be used to help interpret the results of the InCiSE Index, both in terms of the quality of the indicators and for country results.

2.8.1 Quality of indicators

The data quality assessment conducts three checks for each indicator: the availability of metrics, the (non-)use of wider public sector data as a proxy, and the recency of the data. Table 2.8.A summarises the results of these three checks for each of the indicators.

As discussed in sections 2.3 and 2.4 there are four indicators where at least one country is missing all data for the indicator.

Conversely, there is only one indicator (policy making) where all 38 countries have all data available. When it comes to the use of public sector proxy data, there are six indicators where all the data is not a public sector proxy, giving the indicators a maximum proxy data score of 1, and only two indicators (capabilities and digital services) where all the data relates to the civil service and is not public sector proxy which means their maximum proxy score is 0. The recency calculation is a relative assessment where the oldest data (2012) scored 0 and the most recent data (2018) scored 1 – here we see that only one indicator (policy making) is composed solely of 2018 data and again only one indicator (capabilities) is composed solely of 2012 data.

We can also see in Table 2.8.A that there is noticeable variation in the number of

Table 2.8.A   Summary of data quality metadata for the 38 countries of InCiSE 2019

InCiSE indicator	Data availability		Public sector proxy data		Recency of data		Overall DQA score		Countries with max DQA score	Mean DQA score	RAG rating
	Min	Max	Min	Max	Min	Max	Min	Max			
Capabilities	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.33	25	0.244	○
Crisis & risk management	0.00	1.00	0.00	1.00	0.00	0.56	0.00	0.85	18	0.631	◉
Digital services	0.00	1.00	0.00	0.00	0.00	0.74	0.00	0.58	29	0.444	◉
Fiscal & financial management	0.40	1.00	0.50	1.00	0.42	0.67	0.44	0.89	19	0.783	●
HR management	0.60	1.00	0.00	0.44	0.28	0.57	0.29	0.67	34	0.640	◉
Inclusiveness	0.63	1.00	0.20	0.60	0.30	0.57	0.38	0.72	30	0.663	◉
Integrity	0.78	1.00	0.00	0.18	0.43	0.53	0.40	0.57	14	0.526	◉
Openness	0.30	1.00	0.30	1.00	0.25	0.78	0.28	0.93	22	0.818	●
Policy making	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	38	1.000	●
Procurement	0.00	1.00	0.00	0.50	0.00	0.67	0.00	0.72	18	0.513	◉
Regulation	0.35	1.00	0.33	1.00	0.33	0.89	0.34	0.96	34	0.908	●
Tax administration	0.50	1.00	0.33	1.00	0.22	0.56	0.35	0.85	24	0.770	●

● Mean DQA ≥ 0.75   ◉ Mean DQA 0.75-0.25   ○ Mean DQA < 0.25

countries that achieve the maximum overall data quality score for each indicator. For policy making all 38 countries score achieve the maximum score, while for integrity only 14 countries achieve the maximum score. Besides integrity, three other indicators (crisis & risk management, fiscal & financial management, and procurement) have less than 20 countries achieving the maximum score, while three indicators besides policy making have more than 30 countries achieving the maximum score (HR management, inclusiveness, and regulation).

The indicator data quality scores can also be used to create a data-driven red-amber-green (RAG) rating for data quality. Using the mean overall data quality scores for each

indicator from the 38 countries selected for the 2019 edition of InCiSE, a 'green' rating is assigned to those with a score of 0.75 or higher, 'amber' to those with a score between 0.25 and 0.75, and 'red' to those with a score below 0.25.

However, the data quality assessment does not consider the reliability and validity of each indicator's construction and therefore says nothing on how well the indicator represents the concept it is trying to measure. Instead, these data-driven RAG ratings can be combined with a subjective assessment of wider data quality concerns to make an overall assessment of the general 'quality' of each indicator. Table 2.8.B shows the data quality assessment of each indicator

alongside a high-level qualitative assessment of the indicator and a 'final' subjective RAG rating for the indicator.

Five of the indicators have a mean data quality score of 0.75 or higher, earning them an initial 'green' rating. Of these indicators, three retain their green rating after wider considerations of the quality of the indicators are taken into account, meaning that these indicators are considered to provide broad and robust coverage of their respective concepts. Two of the five are demoted from green to amber, reflecting concerns about whether the indicators are sufficiently broad.

Six of the indicators have an initial 'amber' rating. Five of these indicators retain their rating, meaning they may only provide partial coverage of the underlying concept or be heavily reliant on one particular data source or type of data. One of the six is demoted from amber to red, reflecting concerns that the indicator provides limited coverage of the underlying concept.

One indicator has an initial 'red' rating, which is driven largely by its lack of recent data and being solely composed of public sector proxy data. Finally, the social security function, which was included in the 2017 Pilot, is given a 'red' rating following its removal from the 2019 edition of InCiSE due to data quality concerns. This change is discussed further in Chapter 4 and Chapter 6.

2.8.2 Quality of country-level results

Country-level data quality has already been considered to some degree, through the determination of country selection in section 2.3. However, as with the quality of indicators, the results of the data quality assessment can be used to show the relative quality of the selected countries, which can help improve interpretation of the results of the InCiSE Index.

Table 2.8.C presents a detailed overview of the data quality by country. Each country has been given an overall data quality letter "grade" based on its overall data quality score, and for each indicator each country has been given a "RAG" rating.

The overall data quality grades are allocated as follows based on a country's data quality score rounded to 2 decimal places:

A+   for those countries that achieve the highest overall data quality assessment score (i.e. a data quality score of 0.75 when rounded to 2 decimal places)

A   for countries with a data quality score greater than or equal to 0.7 but less than 0.75

B   for countries with a data quality score greater than or equal to 0.65 but less than 0.7

C   for countries with a data quality score greater than or equal to 0.6 but less than 0.65

D   for countries with a data quality score greater than or equal to 0.5 but less than 0.6

For the indicators, a four category "RAG+" rating system is adopted. The data quality scores have been normalised (using min-max normalisation) by indicator:

- A 'green' rating is given to those countries with a normalised indicator data quality score of 1 – the country has the best possible data for this indicator.

- An 'amber' rating is given to those countries with a normalised indicator data quality score of greater than or equal to 0.5 – the country's data quality is at least half as good as the 'best' possible data for that indicator.

- A 'red' rating is given to those countries with a normalised indicator data quality score of less than 0.5 – the country's data quality is less than half as good as the 'best' possible data for that indicator.

- An 'X' rating is given to those countries which have no data at all for that metric – that all of the country's scores for the metrics in that indicator have been imputed.

Table 2.8.C reveals interesting patterns in data quality:

- Six countries are given an "A+" rating – one has full data for all indicators (i.e. all indicator rated 'green'), while the other five have just one indicator where they have an 'amber' rating.

- Eight countries achieve an "A" rating – they have generally good coverage of data but typically have two or three indicators rated 'amber' or 'red", only one country has an indicator where all data for that indicator has been imputed (rated 'grey').

- Seven countries achieve a "B" rating for data quality – these countries have a greater degree of 'amber' and 'red' rated indicators, typically four. All but one country has at least one 'red' rated indicator, one country has one indicator fully imputed while another has two indicators fully imputed.

- Ten countries achieve a "C" rating for data quality – all countries have at least one 'red' rated indicator and eight of the countries have at least one indicator fully imputed.

- Seven countries achieve a "D" rating for data quality – all countries both have at least one indicator fully imputed and one indicator rated 'red', four countries have at least four indicators rated 'red'.

Table 2.8.B Overall quality assessment 'RAG' rating of the 2019 InCiSE indicators

InCiSE indicator	Mean DQA score	Number of metrics	DQA-based RAG rating	High-level assessment of the reliability and validity of the indicator construction	Final RAG rating
Attributes					
Integrity	0.536	17	◉	The indicator has a large number of metrics that give a broad overview of the concept, however it relies heavily on external expert perceptions.	◉
Openness	0.818	10	●	The indicator uses a large number of metrics from a wide range of sources that give a broad overview of the concept.	●
Capabilities	0.244	14	○	While the indicator has a large number of metrics, these are all drawn from a public sector proxy and date between 2012-2015.	○
Inclusiveness	0.663	5	◉	The indicator has only a small number of metrics which only provide a partial picture of performance across the concept.	○
Innovation	–	0	✕	No data available – indicator not measured	✕
Staff engagement	–	0	✕	No data available – indicator not measured	✕
Functions					
Policy making	1.000	8	●	The indicator uses a wide range of metrics that give a broad overview of the concept, however these come from a single source relying on external expert perception.	◉
Fiscal & financial management	0.783	6	●	The indicator contains a number of metrics which appear to give a detailed overview of the concept.	●
Regulation	0.908	9	●	The indicator contains a number of metrics which appear to give a detailed overview of the concept.	●
Crisis & risk management	0.631	13	◉	The indicator contains a wide range of metrics which provide a broad overview of the concept, however one of the two data sources focuses solely on natural disaster risk management.	◉
HR management	0.640	9	◉	The indicator's metrics give an overview of some aspects of the concept, but several metrics are dependent on external perceptions and public sector proxy data.	◉
Procurement	0.513	6	◉	The indicator's metrics give an overview of some aspects of the concept.	◉
Tax administration	0.770	6	●	The indicator has a small number of metrics that give an overview of some aspects of the concept.	◉
Digital services	0.444	13	◉	The indicator relies on a number of metrics from a single source which gives an overview of some aspects of the concept and relies on public sector proxy data.	◉
Social security	–	0	✕	The social security indicator has been depreciated following an in-depth review.	✕
Internal finance	–	0	✕	No data available – indicator not measured	✕
IT for officials	–	0	✕	No data available – indicator not measured	✕

● Mean DQA ≥ 0.75   ◉ Mean DQA 0.75-0.25   ○ Mean DQA < 0.25

Table 2.8.C Data quality scores by indicator and country

Country	Overall data quality		% of metrics available	Data quality of indicator											
	Score	Grade		CAP	CRM	DIG	FFM	HRM	INC	INT	OPN	POL	PRO	REG	TAX
GBR	0.757	A+	100%	●	●	●	●	●	●	●	●	●	●	●	●
ITA	0.755	A+	99%	◉	●	●	●	●	●	●	●	●	●	●	●
POL	0.755	A+	99%	●	●	●	●	●	●	◉	●	●	●	●	●
SWE	0.755	A+	99%	●	●	●	●	●	●	◉	●	●	●	●	●
NOR	0.752	A+	99%	●	●	●	●	●	●	◉	●	●	●	●	●
SVN	0.75	A	99%	●	●	●	●	●	●	●	◉	●	●	●	●
AUT	0.738	A	98%	●	●	●	◉	●	●	◉	●	●	●	●	●
FIN	0.736	A	97%	●	●	●	◉	●	●	○	●	●	●	●	●
ESP	0.733	A	97%	◉	●	●	◉	●	●	◉	◉	●	●	●	●
NLD	0.731	A	98%	●	●	●	◉	●	●	●	●	●	◉	●	●
FRA	0.718	A	97%	◉	●	●	●	●	●	●	●	●	○	●	●
PRT	0.716	A	85%	×	◉	●	●	●	●	◉	●	●	●	●	●
DNK	0.707	A	93%	●	◉	●	◉	●	●	○	◉	●	●	●	◉
DEU	0.701	A	96%	●	●	●	●	●	○	●	●	●	●	●	○
GRC	0.696	B	94%	●	●	●	◉	●	●	◉	●	●	◉	●	○
SVK	0.692	B	93%	●	◉	●	◉	●	●	◉	○	●	●	●	◉
HUN	0.671	B	81%	×	◉	●	●	●	○	●	◉	●	●	●	●
EST	0.669	B	90%	●	○	●	◉	●	○	●	◉	●	●	●	◉
CZE	0.659	B	91%	●	◉	●	●	●	○	◉	●	●	○	●	●
TUR	0.65	C	90%	●	●	●	◉	●	●	◉	●	●	○	●	○
MEX	0.648	C	73%	×	●	×	●	●	●	●	●	●	◉	●	◉
NZL	0.644	C	83%	●	●	×	●	◉	○	◉	●	●	◉	●	●
CHL	0.643	C	79%	●	○	×	●	●	●	●	●	●	◉	●	●
CAN	0.638	C	78%	●	○	×	●	●	●	◉	●	●	◉	●	●
KOR	0.636	C	78%	●	○	×	●	●	●	●	◉	●	◉	●	●
BEL	0.635	C	85%	●	×	●	◉	●	●	●	●	●	○	●	●
LVA	0.628	C	75%	×	○	●	○	●	●	◉	○	●	●	●	●
CHE	0.627	C	79%	×	●	●	◉	●	●	●	○	●	○	●	◉
AUS	0.618	C	71%	×	●	×	●	●	●	◉	●	●	×	●	●
LTU	0.615	C	82%	●	○	●	○	●	●	○	○	●	●	●	●
IRL	0.614	C	84%	●	○	●	◉	●	●	◉	○	●	○	●	●
JPN	0.597	D	75%	●	○	×	●	●	●	◉	●	●	○	●	○
USA	0.579	D	74%	●	○	×	●	●	●	●	●	●	×	◉	◉
ISR	0.578	D	72%	●	○	×	◉	●	●	○	○	●	◉	●	◉
ISL	0.563	D	68%	×	○	●	○	●	●	○	○	●	○	●	◉
ROU	0.529	D	66%	×	◉	●	○	○	○	○	◉	●	●	○	◉
BGR	0.511	D	66%	×	◉	●	○	○	○	○	◉	●	○	○	●
HRV	0.501	D	65%	×	◉	●	○	○	○	○	◉	●	○	○	◉

● High data quality   ◉ Medium data quality   ○ Low data quality   × No data available

2.9 Comparisons over time

The InCiSE project is still in its infancy, and the methodology for the 2019 Index has built substantially on the foundations of the 2017 Pilot – most of the metrics used in the 2017 Pilot have continued to be used in the 2019 edition. Of the 70 metrics in the 2017 Pilot that are directly comparable to the 2019 edition, 33 have since had updates which are incorporated into the model.

In addition to the 70 metrics carried over from the 2017 Pilot, a further 46 metrics have been incorporated into the InCiSE methodology, bringing the total number of metrics for the 2019 model to 116. Most of these additional metrics (30) are from existing sources. Some have been collected multiple times, but some are new and have no previous data collection. Changes are summarised in Chapter 4.

A further consideration for comparisons over time is the need to deal with different reference dates and frequencies of updating. Some data is updated on an annual basis while others are on two-year, three-year, or longer update cycles. For example, the data for capabilities has not been updated since it was first collected in 2012. These differing cycles are the function of a variety of different factors, such as an appreciation of the pace of change within a given topic area or the funding and resourcing of the data producers.

As outlined in section 2.4, the InCiSE model uses imputation methods which use statistical techniques to provide an estimate of a country's missing data. While the imputation is based on predictive methods, it is not a firm prediction of what a given country would have scored, but better understood as indicative. The imputation methods may change between years, and the relationships in the observed data (from

which the imputation is drawn) may also change, limiting the reliability of comparing data imputed in one year with data imputed in another year.

It may also be the case that at one time point a country did not have data for a given metric but then has data at a later time point (or vice versa). This would mean that for one point the metrics would have been imputed. Comparing a score based on 'real' data with one based on imputed estimates is unlikely to be reliable. In addition, as the methodology for InCiSE develops, future versions of the InCiSE Index could adopt back/forward-casting (i.e. using results from different time points) to improve the quality of the imputation methods. This would also make time-series comparison more complicated or less feasible.

Finally, consideration should be given to the changing country composition. The 2017 Pilot covered 31 countries, while the 2019 edition covers 38 countries. As outlined in section 2.5, the data is normalised so that country scores are relative to the group of countries selected. This again means it is not possible to directly compare scores from one edition of InCiSE to another as the scores are related to the specific data range and country set used for that edition.

As a result of these varied challenges, the InCiSE Partners have decided not to include any comparisons between the 2017 Pilot and the 2019 edition of the InCiSE Index. Furthermore, the Partners strongly advise against any direct or indirect comparisons being made beyond references to changes in the underlying source data itself (i.e. before the data is imported into the InCiSE data model, processed, imputed and normalised).

Chapter 3: Methodology of the InCiSE indicators

The following subsections set out the methodology for each of the 12 indicators that make up the 2019 edition of the InCiSE Index. For each indicator this section outlines: the source data; the indicator structure and weighting; the nature and definition of the imported source data and any transformations; the approach to imputation of missing data; and, the rationale for any changes from the 2017 Pilot methodology.

Types of data

The source data for InCiSE comes from a variety of sources which use different methodologies, in this section we have applied the following taxonomy to describe the different types of data sources:

- Subjective data:

    - Public opinion survey – a survey of the opinion/attitudes of the general population/households within a country (e.g. Transparency International's Global Corruption Barometer)

    - Business opinion survey – a survey of the opinion/attitudes of business owners/executives within a country (e.g. the World Economic Forum's Executive Opinion Survey)

    - Expert assessment – a survey/ assessment of a country made by a small number of experts/researchers (e.g the Quality of Government Institute's Expert Opinion Survey)

- Objective data:

    - Analysis of published data – secondary analysis of information/data published by governments

    - Social survey – studies that use scientific social survey methods to collect representative information about the population, but are not opinion surveys (e.g. the OECD's Programme for the International Assessment of Adult Skills)

    - Government assessments – official responses from governments to data collection exercises by international organisations (e.g. OECD surveys)

Each of these types has its strengths and limitations, and some types of data are more appropriate in certain cases than others. The InCiSE model places equal value on these different types of data and does not attempt to make 'quality adjustments', e.g. through weighting, to distinguish between the different types of data.

Critiques of subjective measures can include that they measure perceptions and other 'subjective' positions which may be influenced by considerations beyond just the specific item being measured – e.g. business perceptions of how effective the civil service is at delivering services may be influenced by their perceptions of how business-friendly the government's political programme is. Another critique is through the use of expert assessments, which often rely on a small number of experts/researchers to assess government performance on a given topic or area. However, expert assessments often focus on niche areas which the general public/businesses may not be able to make a judgement about.

Objective data is also not without its own limitations. It can be argued that it is rare for any data to be truly 'objective' even if it is not directly 'subjective'. Even if the data does not aim to measure perceptions or another form of subjective position, it is collected and analysed to fulfil a particular purpose, defined by a particular group of individuals, with a particular agenda. While efforts can be made to minimise biases and particular normative assumptions, in any study there are implicit or explicit subjective decisions made about the collection and analysis of data. The decisions a researcher or analyst makes, such as whether to collect one piece of data over another, which methods of collection and analysis to use, or what to consider in scope or out of scope, are all subjective and therefore will influence the results.

Data sources

Each section lists the data sources used to supply the input data for the InCiSE metrics of each indicator. For ease of reference in each section's tables, the data sources are given an acronym. Figures in square brackets next to a data source indicate the reference year for the data (i.e. the year the data was collected/relates to) rather than the year of publication. A complete reference list of the data sources used for InCiSE is provided in the References chapter.

3.1 Integrity

The integrity indicator is defined as: the extent to which civil servants behave with integrity, make decisions impartially and fairly, and strive to serve both citizens and ministers, and is one of the core values associated with a civil service. The International Civil Service Commission highlights the importance of integrity to the work of the United Nations (UN) common systems staff: "The concept of integrity… embraces all aspects of behaviour of an international civil servant… including … honesty, truthfulness, impartiality and incorruptibility. These qualities are as basic as those of competence and efficiency." (Civil Service Commission, 2002). Numerous studies aiming to establish good governance have utilised similar metrics in their analyses, for instance Muriithi et al. (2015). The inclusion of integrity in the InCiSE is therefore deemed necessary and crucial for the assessment of an effective civil service.

The indicator for integrity is comprised of 17 metrics – an increase of one from the 2017 Pilot edition. A change has also been made in the metric on post-employment cooling-off in the way it has been coded from the source data.

The following sources are used:

- Transparency International's Global Corruption Barometer (GCB) [2017].

- The World Economic Forum's Global Competitiveness Report Executive Opinion Survey (WEF) [2016-2017].

- The University of Gothenburg's Quality of Government Expert Survey (QoG) [2015].

- The OECD's Survey on Managing Conflict of Interest in the Executive Branch and Whistleblower Protection [2014] and Survey on Lobbying Rules and Guidelines [2013] as processed and published in their Government at a Glance 2015 report.

- The Bertelsmann Foundation's Sustainable Governance Indicators (SGI) [2018].

3.1.1 Imputation of missing data

None of the 38 countries selected for the 2019 edition of InCiSE have completely missing data for the integrity metrics. As a result the imputation of missing data for the integrity metrics is based solely on the data within the indicator.

Table 3.1.A Composition of the integrity indicator

Theme	Metric	Source	Type	Public sector proxy	Data transformation	Weighting within indicator		Total (c= A*B)
						In theme (A)	Theme (B)	
Levels/perception of corruption	Officials corrupt	GCB	Public opinion survey	No	None	25%	30%	7.50%
	Business favouritism	WEF	Business opinion survey	Yes	None	25%		7.50%
	Officials embezzle funds	QoG	Expert assessment	Yes	Inverted	25%		7.50%
	Officials grant favours	QoG	Expert assessment	Yes	Inverted	25%		7.50%
Fairness and impartiality	Fair treatment	QoG	Expert assessment	Yes	Inverted	50%	12.5%	6.25%
	Officials act impartially	QoG	Expert assessment	Yes	None	50%		6.25%
Adherence to rules/ procedures	Officials follow rules	QoG	Expert assessment	Yes	None	50%	12.5%	6.25%
	Corruption prevention [new]	SGI	Expert assessment	Yes	None	50%		6.25%
Serving government and citizens	Help citizens	QoG	Expert assessment	Yes	None	33% (1/3)	12.5%	4.16% (1/24)
	Implement ministerial policies	QoG	Expert assessment	Yes	None	33% (1/3)		4.16% (1/24)
	Implement government programme	QoG	Expert assessment	Yes	None	33% (1/3)		4.16% (1/24)
Work ethic	Officials absent	QoG	Expert assessment	Yes	Inverted	50%	12.5%	6.25%
	Strive to be efficient	QoG	Expert assessment	Yes	None	50%		6.25%
Preserving integrity and preventing conflict of interest	Post-employment cooling-off	OECD	Government assessment	No	Composite	25%	20%	5%
	Lobbyist protections	OECD	Government assessment	No	Composite	25%		5%
	Coverage of whistleblower protections	OECD	Government assessment	No	Composite	25%		5%
	Degree of whistleblower protections	QoG	Expert assessment	Yes	Inverted	25%		5%

Table 3.1.B Definition of the integrity metrics

Theme	Metric	Source	Definition of the source metric (e.g. question wording)
Levels/perception of corruption	Officials corrupt	GCB	How many [government officials] do you think are involved in corruption, or haven't you heard enough about them to say? [None, Some of them, Most of them, All of them; % saying none/some]
	Business favouritism	WEF	In your country, to what extent do government officials show favoritism to well-connected firms and individuals when deciding upon policies and contracts? [Rated 1-7; mean score]
	Officials embezzle funds	QoG	Public sector employees steal, embezzle or misappropriate public funds or other state resources for personal or family use [Rated 1-7; mean score]
	Officials grant favours	QoG	Public sector employees grant favors in exchange for bribes, kickbacks or other material inducements [Rated 1-7; mean score]
Fairness and impartiality	Fair treatment	QoG	When deciding how to implement policies in individual cases, public sector employees treat some groups in society unfairly [Rated 1-7; mean score]
	Officials act impartially	QoG	Generally speaking, how often would you say that public sector employees today, in your chosen country, act impartially when deciding how to implement a policy in an individual case? [Rated 1-7; mean score]
Adherence to rules/ procedures	Officials follow rules	QoG	Public sector employees strive to follow rules [Rated 1-7; mean score]
	Corruption prevention [new]	SGI	To what extent are public officeholders prevented from abusing their position for private interests? [Rated 1 to 10; mean score]
Serving government and citizens	Help citizens	QoG	Public sector employees strive to help citizens [Rated 1-7; mean score]
	Implement ministerial policies	QoG	Public sector employees strive to implement the policies decided upon by the top political leadership [Rated 1-7; mean score]
	Implement government programme	QoG	Public sector employees strive to fulfill the ideology of the party/parties in government [Rated 1-7; mean score]
Preserving integrity and preventing conflict of interest	Post-employment cooling-off	OECD	Public employees are absent from work without permission [Rated 1-7; mean score]
	Lobbyist protections	OECD	Public sector employees strive to be efficient [Rated 1-7; mean score]
	Coverage of whistleblower protections	OECD	Existence of post-employment cooling-off period for senior civil servants and all other civil servants
	Degree of whistleblower protections	QoG	Three binary (yes/no) variables on lobbyist protections

3.1.2 Changes from the 2017 Pilot

There is one proposed change to the structure of the metrics used in the calculation of the integrity indicator: the inclusion of a measure from the Bertelsmann Foundation's Sustainable Governance Indicators on corruption prevention.

A further change from the 2017 pilot methodology has been implemented in the processing of the OECD's data on post-employment cooling-off periods. The model now simply codes whether post-employment cooling-off periods and now ignores whether compensation is paid during this period.

The OECD source data provides information on whether post-employment cooling-off periods exist for both senior civil servants and other civil servants, and also includes information on whether a compensation period is paid during that period.

These data are combined by the InCiSE model into a single scale, outlined below. In the 2017 Pilot, this scale creates the normative conditions that a post-employment cooling-off period with compensation for both groups of civil servants is "best" and no cooling-off period is "worst".

Further examination of the data, as reported by the OECD, showed that only a limited number of officials in only a small number of countries received paid compensation during a cooling off period and that there was noticeable variation in how this was decided by country. This limited usage of post-employment compensation and high variability in its nature suggests that it may not be appropriate to code in the provision of post-employment compensation as normative "best" practice in the calculation of the integrity indicator.

Table 3.1.C Coding of post-employment cooling-off in the 2017 Pilot edition of InCiSE

Senior civil servants		Other civil servants		2017 post-employment scale value
Cooling-off period?	With compensation?	Cooling-off period?	With compensation?	
Yes	Yes	Yes	Yes	4
Yes	Yes	Yes	No	3
Yes	No	Yes	Yes	
Yes	Yes	No	N/A	2
No	N/A	Yes	Yes	
Yes	No	Yes	No	
Yes	No	No	N/A	1
No	N/A	Yes	No	
No	N/A	No	N/A	0

> *During the cooling off period, only some categories of public officials in Austria, Israel, Norway, Portugal and Spain receive compensation. For instance, in Spain, public officials receive 80% of their basic salaries as compensation and in Norway, compensation is awarded only for prohibitions on taking up a specific appointment, the level of which is equivalent to the salary received at the time of the public official left public office"*
> OECD (2015) Government at a Glance 2015, p116

Therefore, for the 2019 edition, InCiSE has adopted a new scale that measures only the existence of post-employment cooling-off periods for senior civil servants and other civil servants, ignoring the use/existence of compensation. The highest score will be awarded for those countries that have a cooling-off period for both groups of civil servants, the lowest score for those that do not have a cooling-off period for either group, while an intermediate score will be given to those countries that have a cooling-off period for one group but not the other – with cooling-off periods for senior civil servants preferred to those for non-senior civil servants.

Table 3.1.D   Coding of post-employment cooling-off in the 2019 edition of InCiSE

Does a post-employment cooling-off period exist?		2019 post-employment scale value
Senior civil servants	Other civil servants	
Yes	Yes	3
Yes	No	2
No	Yes	1
No	No	0

3.2 Openness

The openness indicator is defined as: the regular practice and degree of consultation with citizens to help guide the decisions we make and extent of transparency in our decision-making. It is included in the index because the need for transparency within a civil service is imperative for the public to trust and feel empowered to hold the government accountable for their actions, whilst at the same time reducing corruption. The World Bank (2017) notes that "transparency initiatives [are] an important first step toward increasing accountability". The UN also outlines the need for transparency and accountability in governance: "[this] implies a proactive effort to make information accessible to citizens" and it is "one indicator of a government that is citizen-focused and service-oriented". (United Nations, 1999). Graham et al. also refer to the United Nations Development Program's five principles of good governance, in which transparency is identified as a key characteristic.

This indicator is comprised of 10 metrics, an increase of one from the 2017 Pilot edition of InCiSE. The data sources for the openness indicator are:

- The open government domain of the World Justice Project's Rule of Law Index (RLI) [2017].

- The United Nations' E-Participation Index (UN) [2018].

- Bertelsmann Stiftung's Sustainable Governance Indicators (SGI) [2018].

- The World Wide Web Foundation's Open Data Barometer (ODB) [2016].

- Open Knowledge International's Global Open Data Index (OKI) [2016].

- The OECD's Open, Useful, Reusable (OUR) Government Data Index (OECD) [2016].

3.2.1 Imputation of missing data

None of the 38 countries selected for the 2019 edition of InCiSE have completely missing data for the openness metrics. As a result the imputation of missing data for the openness metrics is based solely on the data within the indicator.

3.2.2 Changes from the 2017 Pilot

Compared to the 2017 Pilot, an additional metric from the Bertelsmann Sustainable Governance Indicators on access to information has been identified and added to the indicator.

Table 3.2.A Composition of the openness indicator

Theme	Metric	Source	Type	Public sector proxy	Data transformation	Weighting within indicator		Total (c= A*B)
						In theme (A)	Theme (B)	
The degree and quality of societal consultation	Civic Participation	RLI	Mix of public opinion survey and expert assessment	No	None	33% (1/3)	37.5%	12.5%
	E-Government Engagement	UN	Expert assessment	No	None	33% (1/3)		12.5%
	Negotiating Public Support	SGI	Expert assessment	No	None	33% (1/3)		12.5%
Existence of complaint mechanism	Complaint Mechanisms	RLI	Mix of public opinion survey and expert assessment	No	None	100%	12.5%	12.5%
Government data availability and usability	Open Data Practice and Impact	ODB	Mix of expert assessments and analysis of published data	No	None	33% (1/3)	30%	10%
	Government Datasets Openness	OKI	Analysis of published data	No	None	33% (1/3)		10%
	Data Availability and Government Support	OECD	Expert assessment	No	None	33% (1/3)		10%
Right to information (e.g. freedom of information)	Right to Information	RLI	Mix of public opinion survey and expert assessment	No	None	50%	10%	5%
	Access to Government Information *[new]*	SGI	Expert assessment	No	None	50%		5%
Publicised laws	Publicised laws	RLI	Mix of public opinion survey and expert assessment	No	None	100%	10%	10%

Table 3.2.B Definition of the openness metrics

Theme	Metric	Source	Definition of the source metric (e.g. question wording)
The degree and quality of societal consultation	Civic Participation	RLI	Average of 31 opinion/assessment items measuring the effectiveness of civic participation and feedback mechanisms. [Normalised index ranging from 0 to 1]
	E-Government Engagement	UN	Reviews the quality and usefulness of e-government programs for the purpose of engaging people in public policy-making and implementation. [Normalised index ranging from 0 to 1]
	Negotiating Public Support	SGI	To what extent does the government consult with societal actors to support its policy?[Rated 1 to 10]
Existence of complaint mechanism	Complaint Mechanisms	RLI	Average of 5 opinion/assessment items measuring whether people are able to complain about public services/officials, how government responds. [Normalised index ranging from 0 to 1]
Government data availability and usability	Open Data Practice and Impact	ODB	How governments are publishing and using open data for accountability, innovation and social impact. [Index ranging from 0 to 100]
	Government Datasets Openness	OKI	Assessment of the 'openness' of government datasets. [Index ranging from 0 to 100]
	Data Availability and Government Support	OECD	Assesses government's open data efforts across the domains of openness, usefulness and reusability. [Index ranging from 0 to 100]
Right to information (e.g. freedom of information)	Right to Information	RLI	Average of 25 opinion/assessment items on how government handles requests for information. [Normalised index ranging from 0 to 1]
	Access to Government Information *[new]*	SGI	To what extent can citizens obtain official information? [Rated 1 to 10]
Publicised laws	Publicised laws	RLI	Average of 17 opinion/assessment items on how whether laws and legal rights are publicly available (including in plain language). [Normalised index ranging from 0 to 1]

3.3 Capabilities

The capabilities indicator is defined as: the extent to which the workforce has the right mix of skills. The need for a variety of certain strong skills is vital for the successful operation of any organisation, civil services included. The standards for good governance set out by the Office for Public Management (OPM) and the Chartered Institute of Public Finance and Accountability (CIPFA) include leadership as a core skill. It goes on to list necessary skills as "the ability to scrutinise and challenge information… including skills in financial management and the ability to recognise when outside expert advice is needed" (2004). Fukuyama acknowledges the importance of educational attainment of civil servants: "another critical measure of capacity is the level of education and professionalisation of government officials", along with the importance of digital capability: "what level of technical expertise they are required to possess" (2013).

The capabilities indicator is composed of 14 metrics from the OECD's Programme for the International Assessment of Adult Competencies (referred to as PIAAC from this point onwards), this is an increase of 10 metrics from the 2017 Pilot.

PIAAC is a scientific assessment of competencies in adults, modelled on the OECD's successful Programme for International Student Assessment (PISA) that measures the competencies of school-aged children around the world. Data for 25 countries was collected over 2011-12, and data for nine countries was collected over 2014-15. Of these, 31 countries have published microdata available for analysis.

The results from PIAAC are not published in a form that allows for direct import of the relevant data for InCiSE. Instead the data must be calculated from the individual respondent-level microdata published by the OECD. The microdata is analysed to produce results for those defined as currently working in the "public administration" sector of the International Standard Industrial Classification. This is wider than just the civil service and includes other forms of public administration, such as sub-national and local government, but excludes functions such as healthcare, education and transport which may or may not be part of the public sector depending on country.

3.3.1 Imputation of missing data

Of the 38 countries selected for the 2019 edition of InCiSE, 10 countries do not have data for the capabilities metrics. As there are countries where data is missing for all metrics the imputation of the capabilities indicator requires a data point from outside the indicator. The 2017 edition of InCiSE used data from the HR Management indicator on applicant skills and whether a country was an EU member. For the 2019 edition, the applicant skills metric from the HR management indicator is retained, but EU membership is removed. One of the metrics within the indicator is the level of tertiary educational attainment. There are a number of sources for estimates of tertiary educational attainment in the general adult population of most countries. Therefore, InCiSE also uses UNESCO data on educational attainment to impute missing data for the capabilities indicator.
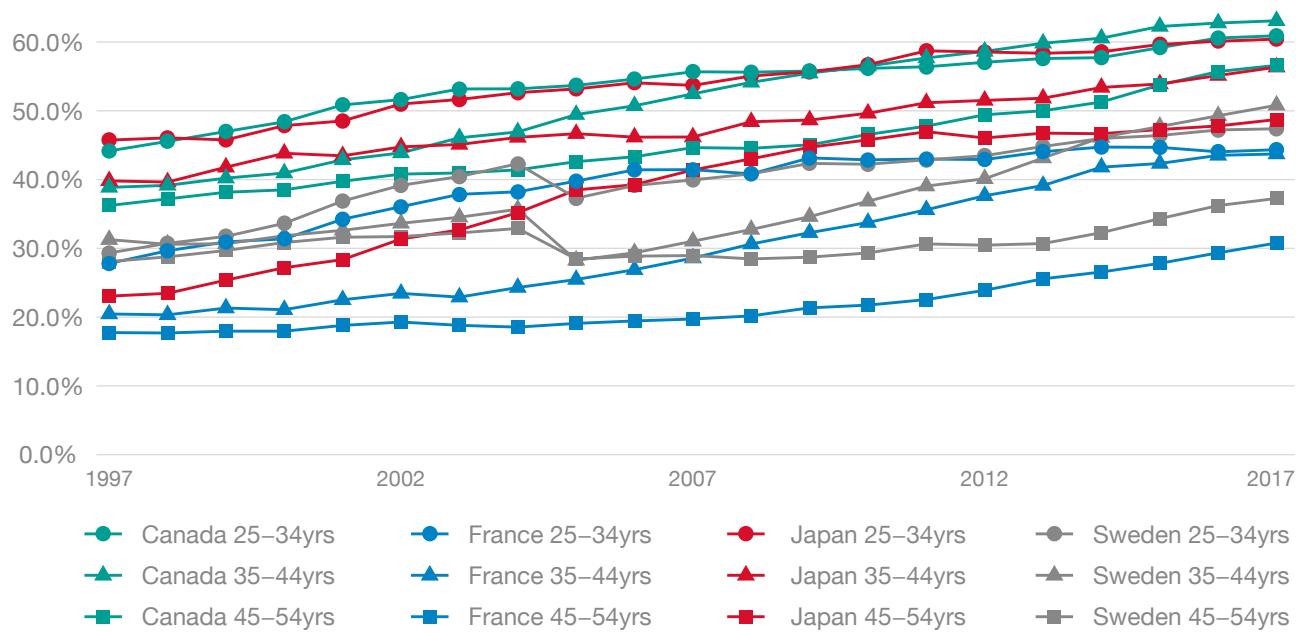
Table 3.3.A Composition of the capabilities indicator

Theme	Metric	Source	Type	Public sector proxy	Data transformation	Weighting within indicator		Total (c= A*B)
						In theme (A)	Theme (B)	
Core capability (educational attainment and skill levels)	Literacy capability	PIAAC	Social survey	Yes	None	25%	25%	6.25%
	Numeracy capability	PIAAC	Social survey	Yes	None	25%		6.25%
	Problem-solving capability	PIAAC	Social survey	Yes	None	25%		6.25%
	Educational attainment	PIAAC	Social survey	Yes	None	25%		6.25%
Use of core skills at work	ICT use at work [new]	PIAAC	Social survey	Yes	None	25%	25%	6.25%
	Numeracy at work [new]	PIAAC	Social survey	Yes	None	25%		6.25%
	Reading at work [new]	PIAAC	Social survey	Yes	None	25%		6.25%
	Writing at work [new]	PIAAC	Social survey	Yes	None	25%		6.25%
Organisational skills	Influencing at work [new]	PIAAC	Social survey	Yes	None	33% (1/3)	25%	8.33% (1/12)
	Planning at work [new]	PIAAC	Social survey	Yes	None	33% (1/3)		8.33% (1/12)
	Task discretion [new]	PIAAC	Social survey	Yes	None	33% (1/3)		8.33% (1/12)
Learning and development	Learning at work [new]	PIAAC	Social survey	Yes	None	33% (1/3)	25%	8.33% (1/12)
	Openness to learning [new]	PIAAC	Social survey	Yes	None	33% (1/3)		8.33% (1/12)
	Learning in the past year [new]	PIAAC	Social survey	Yes	None	33% (1/3)		8.33% (1/12)

Table 3.3.B Definition of the capabilities metrics

Theme	Metric	Source	Definition of the source metric (e.g. question wording)
Core capability (educational attainment and skill levels)	Literacy capability	PIAAC	Percentage of adults scoring high (level 4/5) in literacy.
	Numeracy capability	PIAAC	Percentage of adults scoring high (level 4/5) in numeracy.
	Problem-solving capability	PIAAC	Percentage of adults scoring at level 3 in problem-solving in a technology-rich environment.
	Educational attainment	PIAAC	Percentage of public sector workforce with tertiary education.
Use of core skills at work	ICT use at work [new]	PIAAC	Index of use of ICT skills at work (email, internet, spreadsheets, word processing)
	Numeracy at work [new]	PIAAC	Index of use of numeracy skills at work (calculating costs/numbers, charts/graphs, maths/statistics)
	Reading at work [new]	PIAAC	Index of use of reading skills at work (correspondence, media, reference/ academic material, diagrams)
	Writing at work [new]	PIAAC	Index of use of writing skills at work (correspondence, media, reports, filling in forms)
Organisational skills	Influencing at work [new]	PIAAC	Index of use of influencing skills at work (giving presentations/advice, instructing others, negotiation)
	Planning at work [new]	PIAAC	Index of use of planning skills at work (planning own and others' activities, organising own time)
	Task discretion [new]	PIAAC	Index of task discretion at work (choice in sequence of tasks, how to do work, rate, working hours)
Learning and development	Learning at work [new]	PIAAC	Index of learning in the workplace (learn from colleagues, learn by doing, need to keep up-to-date)
	Openness to learning [new]	PIAAC	Index of readiness to learn (interest in learning, like solving problems, seek out new information)
	Learning in the past year [new]	PIAAC	Percentage who have participated in formal or informal learning for job-related reasons in past 12 months

Figure 3.1   Tertiary education levels of adults over 25, by age group, in selected countries

Legend:
- Canada 25–34yrs
- Canada 35–44yrs
- Canada 45–54yrs
- France 25–34yrs
- France 35–44yrs
- France 45–54yrs
- Japan 25–34yrs
- Japan 35–44yrs
- Japan 45–54yrs
- Sweden 25–34yrs
- Sweden 35–44yrs
- Sweden 45–54yrs

Source: OECD (2018), Population with tertiary education (indicator). doi: 10.1787/0b8f90e9-en (Accessed on 15 August 2018)

3.3.2 Changes from the 2017 Pilot

The capabilities indicator published in the 2019 edition of InCiSE has had a number of changes which improve its quality compared to the data published in the 2017 Pilot. These include additional metrics, change in how data is extracted, updated coding of educational attainment, and changes to imputation. While these do not change the recency of the data, they improve the overall quality of the information. The OECD intends to update PIAAC every decade, as annual change in the skill level of the adult population does not change rapidly – a general principle in education research is that educational attainment is broadly fixed after young adulthood.[3] Figure 3.1, shows how the overall proportion of tertiary educational attainment has evolved for different age groups since 1997 in four countries, the average annual change is 0.9 percentage points.

Additional metrics

In examining the PIAAC dataset, a number of additional metrics that complement the metrics used in the pilot provide a richer picture of capabilities in the public administration workforce.

The pilot metrics give a broad overview of employee capability, looking at overall levels of core skills (literacy, numeracy and problem solving) and tertiary educational attainment. The additional metrics complement this by providing for measurement of the use of core

---

3   Lutz et al (2007) and Gujon et al (2016) utilise this principle to develop "back-projections" of educational attainment, and hold a general assumption that 'transition' to different levels of education tend to be limited after the age of 34.

skills at work (ICT, numeracy, reading and writing). They also cover more complex skills, including influencing others, planning, and task management. Finally, they also include metrics relating to learning and development, i.e. whether individuals learn at work, their overall attitude to learning, and whether they have participated in learning for work-related purposes (either formally or informally). Together these metrics provide a more detailed picture of the skills and capabilities of the workforce.

Using the public administration industrial sector

The pilot edition of InCiSE used data for all adults currently employed by a public sector organisation. Further investigations of the raw data in PIAAC indicated that there was a sufficient sample size in most countries (n>100) to generate a reliable estimate for the "public administration" industry sector.[4] This was further limited to those who said they worked for a public sector organisation.

There is a considerable difference between countries with regard to whether someone is a public sector worker. This is in part due to the political choices about what is or isn't delivered by the public sector. For example, in the United Kingdom the vast majority of healthcare workers will be public sector employees, while in the United States the vast majority of healthcare workers will be private sector employees. In contrast, this difference is likely to be much reduced for the "public administration" industry sector, as it will not include sectors such as healthcare, education or competitive market economic sectors. Therefore, while the sample size for the "public administration" subset will be lower, it is likely

to be a more appropriate comparator group across countries than using the large "public sector" basis.

Further details on the structure of the activities included in the "public administration" industrial sector can be found in the UN's registry of Statistical Classifications (UNSD 2018).

Updated coding of tertiary education

In reviewing the way that results are extracted from PIAAC's raw data files, an improvement was identified in the way tertiary education is coded. The pilot edition of InCiSE used data from a variable included for legacy comparisons with previous international assessments of adult competencies based on type of institution attended. This year, InCiSE 2018 will use a more accurate method based on the highest level of qualification achieved.

Updating the approach to imputation

In the pilot edition of InCiSE, missing data issues were handled by examining the relationship of the metrics from PIAAC with metrics from the other indicators in InCiSE (as PIAAC is the only data source for the capabilities indicator). The most suitable predictors observed in the dataset were the applicant skills metric from the HR management indicator and whether a country was an EU member. As described above, the imputation for the 2019 edition has changed the methodology to remove the EU membership criteria and include the tertiary education level of the general population in the external imputation data. This provides a closer intellectual link with the indicator's construct.

4   Sample sizes for the public administration industry sector (limited to declared public sector workers) range from 83 to 1,562. The minimum and maximum are both noticeable outliers: ignoring these, the sample sizes range from 144-446. The only country with a sample less than 100 (Russian Federation) had similar standard errors to those of other countries and therefore was retained in the data extracted from PIAAC.

3.4 Inclusiveness

The inclusiveness indicator is defined as: the extent to which the civil service is representative of the citizens it serves. A model civil service should be representative of the public it stands to serve, and therefore institutions must be inclusive in nature. In their Post-2015 Millennium Development Goal reflections, the OECD (n.d.) outlines the greater success felt by inclusive public bodies: "Inclusive governments and an active civil society put forward more responsive, equitable policies" and that these "build trust in government and help create… public services that are better suited to diverse needs". The guiding principles to the international civil service, set out by the International Civil Service Commission, support the claim that civil servants must "respect the dignity, worth and equality of all people" and have: "a willingness to work without bias with persons of all nationalities, religions and cultures" (2002). The OECD Government at a Glance (2015a) report states that "a more representative public administration can better access previously overlooked knowledge, networks and perspectives for improved policy development and implementation". The same report also points out that the opinion on the groups in need of representation in public administration has widened "and now includes a range of dimensions such as women; racial, ethnic, and religious minorities; the poor; the elderly; the disabled; and other minority groups such as indigenous populations". A paper by the Office for Public Management (OPM) and the Chartered Institute of Public Finance and Accountability (CIPFA) highlights the potential benefits of this view: "Public trust and confidence in governance will increase if governance … [is] done by a diverse group of people who reflect the community" (2004).

The inclusiveness indicator is comprised of five metrics, and is unchanged in structure from the 2017 Pilot. It uses the following source data:

- OECD data on the central government share of women in the central government and in top management positions [2016], as processed and published in their Government at a Glance 2017 report.

- The University of Gothenburg's Quality of Government Expert Survey (QoG) [2015].

- Figures on women's representation in the government workforce are compared to data from the International Labour Organisation on the composition to calculate the difference between government and the workforce as a whole (ILO) [2015].

3.4.1 Imputation of missing data

None of the 38 countries selected for the 2019 edition of InCiSE have completely missing data for the inclusiveness metrics. As a result the imputation of missing data for the inclusiveness metrics is based solely on the data within the indicator.

3.4.2 Changes from the 2017 Pilot

There are no changes in the structure of the indicator from the 2017 Pilot.

Table 3.4.A Composition of the inclusiveness indicator

Theme	Metric	Source	Type	Public sector proxy	Data transformation	Weighting within indicator		
						In theme (A)	Theme (B)	Total (c= A*B)
Gender	Central government share	OECD	Government assessment	No	Composite	25%	75%	18.75%
	Public sector share	QoG	Expert assessment	Yes	Composite	25%		18.75%
	Top management share	OECD	Government assessment	No	Composite	25%		18.75%
	Senior government share	QoG	Expert assessment	Yes	Composite	25%		18.75%
Ethnic minorities	Ethnic and religious group representation	QoG	Expert assessment	No	None	100%	25%	25%

Table 3.4.B Definition of the inclusiveness metrics

Theme	Metric	Source	Definition of the source metric (e.g. question wording)
Gender	Central government share	OECD	Share of total central government employment filled by women
	Public sector share	QoG	Thinking about the country for which you have chosen to submit your answers, what is the proportion of women among public sector employees?
	Top management share	OECD	Share of women in top management positions in central government
	Senior government share	QoG	Thinking about the country for which you have chosen to submit your answers, what is the proportion of women among senior positions in central government?
Ethnic minorities	Ethnic and religious group representation	QoG	Key ethnic and religious groups in society are proportionally represented among public sector employees [Rated 1-7; mean score]

3.5 Policy making

The policy making indicator is defined as: the quality of the policy making process, including how policy is developed and coordinated across government and how policy is monitored during implementation. Policy making remains a central role of a civil service and the quality of evidence and appraisal are central to the success of policy. Kaufmann et al. (1999) outline three functions of good governance, including "the capacity of government to effectively formulate and implement sound policies". Policymakers need to "receive rigorous analyses of comprehensive background information and evidence, and of the options for actions" according to the Office for Public Management (OPM) and the Chartered Institute of Public Finance and Accountability (CIPFA) (2004). This paper also advises that "good quality information and clear, objective advice can significantly reduce the risk of taking decisions that fail to achieve their objectives or have serious unintended consequences".

The indicator is comprised of eight metrics, and the structure is unchanged from the 2017 Pilot edition of InCiSE. The policy making indicator uses a single source, the Bertelsmann Stiftung's Sustainable Government Indicators (SGI), an expert assessment of the performance of government in EU and OECD countries. The data for the 2019 edition of InCiSE use the 2018 edition of the SGIs.

3.5.1 Imputation of missing data

All 38 countries selected for the 2019 edition of InCiSE have data for all the metrics in the policy making indicator. Therefore, no approach to imputation is needed.

3.5.2 Changes from the 2017 Pilot

The policy making indicator is unchanged from the 2017 Pilot edition.

Table 3.5.A Composition of the policy making indicator

Theme	Metric	Source	Type	Public sector proxy	Data transformation	Weighting within indicator		Total (c= A*B)
						In theme (A)	Theme (B)	
Quality of policy advice	Scholarly advice	SGI	Expert assessment	No	None	50%	25%	12.5%
	Government office expertise	SGI	Expert assessment	No	None	50%		12.5%
Strategic planning	Strategic planning	SGI	Expert assessment	No	None	100%	25%	25%
Policy coordination	Ministerial coordination	SGI	Expert assessment	No	None	50%	25%	12.5%
	Line ministries involve the centre	SGI	Expert assessment	No	None	50%		12.5%
Policy monitoring	Monitoring of ministries	SGI	Expert assessment	No	None	33% (1/3)	25%	8.33% (1/12)
	Monitoring of agencies	SGI	Expert assessment	No	None	33% (1/3)		8.33% (1/12)
	National standards	SGI	Expert assessment	No	None	33% (1/3)		8.33% (1/12)

Table 3.5.B Definition of the policy making metrics

Theme	Metric	Source	Definition of the source metric (e.g. question wording)
Quality of policy advice	Scholarly advice	SGI	How influential are non-governmental academic experts for government decision- making? [Rated 1-10; mean score]
	Government office expertise	SGI	Does the government office / prime minister's office have the expertise to evaluate ministerial draft bills substantively? [Rated 1-10; mean score]
Strategic planning	Strategic planning	SGI	How much influence do strategic planning units and bodies have on government decision-making? [Rated 1-10; mean score]
Policy coordination	Ministerial coordination	SGI	How effectively do ministry officials / civil servants coordinate policy proposals? [Rated 1-10; mean score]
	Line ministries involve the centre	SGI	To what extent do line ministries involve the government office / prime minister's office in the preparation of policy proposals? [Rated 1-10; mean score]
Policy monitoring	Monitoring of ministries	SGI	How effectively does the government office/prime minister's office monitor line ministry activities with regard to implementation? [Rated 1-10; mean score]
	Monitoring of agencies	SGI	How effectively do federal and subnational ministries monitor the activities of bureaucracies / executive agencies with regard to implementation? [Rated 1-10; mean score]
	National standards	SGI	To what extent does central government ensure that subnational self-governments realize national standards of public services? [Rated 1-10; mean score]

3.6 Fiscal and financial management

The fiscal and financial management indicator is defined as: The quality of the budgeting process and the extent to which spending decisions are informed through economic appraisal and evaluation. It is an important measure of every system of public administration. The Indicator of the Strength of Public Management Systems (ISPMS) from the World Bank state "Public sector management arrangements must also encourage fiscal and institutional sustainability as less tangible but equally critical outcomes" and "Reforms of budgetary and financial management systems… are often crucial for development outcomes" (2012). Holt and Manning (2014) also consider that "public administration practitioners break down the functioning of the central agencies into five management systems", including fiscal and financial management which is made up of: "planning and budgeting; financial management; and accounting, fiscal reporting and audit". The OECD's recommendation paper on budgetary governance (2015b) also sets out ten principles for good budgetary governance which include "ensur[ing] that performance, evaluation, and value for money are integral to the budget process… [and] …manag[ing] budgets within clear, credible and predictable limits for fiscal policy".

The fiscal and financial management indicator is made up of six metrics, an increase of three from the 2017 Pilot. The sources for the indicator are:

- The OECD's 'medium-term budgeting index' [2012] and 'performance budgeting index' [2016].

- The World Economic Forum's Global Competitiveness Index (WEF) [2016-2017].

- World Bank Financial Management Information Systems & Open Budget Data (WB) [2017].

- International Budget Partnership' s Open Budget Survey (IBP) [2017].

3.6.1 Imputation of missing data
None of the 38 countries selected for the 2019 edition of InCiSE have completely missing data for the fiscal and financial management metrics. As a result the imputation of missing data for the fiscal and financial management metrics is based solely on the data within the indicator.

3.6.2 Changes from the 2017 Pilot
The fiscal and financial management indicator has seen the introduction of three new data points to increase the scope and robustness of the indicator. These include a metric on the publication of medium-term budgeting data from the World Bank into the theme of the same name and two new metrics under the economic appraisal and evaluation theme: two data points measuring the extent of external scrutiny or audit and two data points measuring the extent of transparency based on the publication of budgetary reports.

Table 3.6.A Composition of the fiscal and financial management indicator

Theme	Metric	Source	Type	Public sector proxy	Data transformation	Weighting within indicator		
						In theme (A)	Theme (B)	Total (c= A*B)
Effectiveness of public spending	Efficiency of public spending	WEF	Business opinion survey	No	None	100%	10%	10%
Transparency of public spending	Openness of budget process *[new]*	IBP	Expert assessment	No	None	40%	50%	20%
	Openness of financial information *[new]*	WB	Analysis of published data and expert assessment	No	None	40%		20%
	Published public finance data *[new]*	WB	Analysis of published data and expert assessment	No	Composite	20%		10%
Budgeting practice	Medium-term Budgeting	OECD-MT	Government assessment	No	None	50%	40%	20%
	Performance Budgeting	OECD-PB	Government assessment	No	None	50%		20%

Table 3.6.B Definition of the fiscal and financial management metrics

Theme	Metric	Source	Definition of the source metric (e.g. question wording)
Effectiveness of public spending	Efficiency of public spending	WEF	How would you rate the composition of public spending in your country? [Rated 1 (extremely wasteful) to 7 (highly efficient in providing the necessary goods and services)]
Transparency of public spending	Openness of budget process *[new]*	IBP	Index of the openness of budget process [Composite score ranging from 0 to 100]
	Openness of financial information *[new]*	WB	Composite score derived from assessments of what information about public finances is published [Composite score ranging from 0 to 100]
	Published public finance data *[new]*	WB	Categories of public expenditure/revenue published
Budgeting practice	Medium-term Budgeting	OECD-MT	Use of a medium-term perspective in the budget process [Index ranging from 0 to 1]
	Performance Budgeting	OECD-PB	Use of performance budgeting at the central level of government [Index ranging from 0 to 1]

3.7 Regulation

The regulation indicator is defined as: the extent and quality of regulatory impact assessments and the degree of stakeholder engagement involved in them. The appropriate appraisal and evaluation of regulatory changes accompanied by sufficient stakeholder engagement is crucial to ensuring that any introductions are fully considered and fair, involving various stakeholders. This scrutiny is endorsed by many; the OECD for instance, "recognis[es] that regulations are one of the key levers by which governments act to promote economic prosperity, enhance welfare and pursue the public interest", and that "well designed regulations can generate significant social and economic benefits which outweigh the costs of regulation, and contribute to social well-being" (2012). The International Monetary Fund (IMF) acknowledges the importance of regulatory frameworks to successful governance: "From the perspective of the IMF, countries with good governance have strong legal and regulatory frameworks in place" (2016). Additionally, in promoting best practice, "[the] Regulatory Impact Analysis (RIA) is a multiple stakeholder assessment of the economic, environmental and social impact of regulations. The OECD and European Union have strongly promoted this evidence-based approach towards legislation" (Bovaird and Löffler, 2003).

The regulation indicator is comprised of nine metrics, an increase of three from the 2017 Pilot. It uses the following sources:

- The OECD's Indicators of Regulatory Policy and Governance (OECD) [2017].

- The Bertelsmann Foundation's Sustainable Governance Indicators (SGI) [2018].

3.7.1 Imputation of missing data

None of the 38 countries selected for the 2019 edition of InCiSE have completely missing data for the regulation metrics. As a result the imputation of missing data for the regulation metrics is based solely on the data within the indicator.

3.7.2 Changes from the 2017 Pilot

The regulation indicator has had three additional metrics added from the Bertelsmann Foundation's Sustainable Governance Indicators on the use and quality of regulatory impact assessments (RIA), and whether RIAs include sustainability checks.

Table 3.7.A Composition of the regulation indicator

Theme	Metric	Source	Type	Public sector proxy	Data transformation	Weighting within indicator		
						In theme (A)	Theme (B)	Total (c= A*B)
Use of appraisal and evaluation	Use of impact assessments (n)	SGI	Expert assessment	No	None	20%	50%	10%
	Impact assessment for primary law	OECD	Government assessment	No	None	20%		10%
	Impact assessment for secondary law	OECD	Government assessment	No	None	20%		10%
	Evaluation for primary law	OECD	Government assessment	No	None	20%		10%
	Evaluation for secondary law	OECD	Government assessment	No	None	20%		10%
Stakeholder engagement	Stakeholder engagement for primary law	OECD	Government assessment	No	None	50%	25%	12.5%
	Stakeholder engagement for secondary law	OECD	Government assessment	No	None	50%		12.5%
Nature of impact assessment	Quality of RIA (n)	SGI	Expert assessment	No	None	50%	25%	12.5%
	Sustainability check (n)	SGI	Expert assessment	No	None	50%		12.5%

Table 3.7.B Definition of the regulation metrics

Theme	Metric	Source	Definition of the source metric (e.g. question wording)
Use of appraisal and evaluation	Use of impact assessments (n)	SGI	To what extent does the government assess the potential impacts of existing and prepared legal acts (regulatory impact assessments, RIA)? [rated 1-10; mean score]
	Impact assessment for primary law	OECD	OECD composite of government regulatory governance practices
	Impact assessment for secondary law	OECD	OECD composite of government assessments of the use of impact assessment
	Evaluation for primary law	OECD	OECD composite of government assessments of the use of impact assessment
	Evaluation for secondary law	OECD	OECD composite of government assessments of the use of impact assessment
Stakeholder engagement	Stakeholder engagement for primary law	OECD	OECD composite of government assessments of the use of impact assessment
	Stakeholder engagement for secondary law	OECD	OECD composite of government assessments of the use of impact assessment
Nature of impact assessment	Quality of RIA (n)	SGI	Does the RIA process ensure participation, transparency and quality evaluation? [rated 1-10; mean score]
	Sustainability check (n)	SGI	Does the government conduct effective sustainability checks within the framework of RIA? [rated 1-10; mean score]

3.8 Crisis and risk management

The crisis and risk management indicator is defined as: the effectiveness with which the government engages the whole of society to better assess, prevent, respond to and recover from the effects of extreme events. The OECD Strategic Crisis Management report highlights crisis management as central to government's role and a "fundamental element of good governance" (Baubion, 2013). Studies have shown that credibility and trust in governments to deal with crises is vital both to reassure and encourage support from the private sector and general public, as outlined by Christensen et al. (2011).

The crisis and risk management indicator is made up of 13 metrics. This is an increase of four from the 2017 Pilot, however it has been restructured to allow for the inclusion of a new data source, with eight metrics continuing from the 2017 Pilot and five new metrics. The data for the indicator comes from:

- The United Nation's Hyogo Framework for Action monitoring reports [2015].

- The OECD's Survey on the Governance of Critical Risk [2016].

Both the Hyogo Framework monitoring reports and the OECD survey are largely composed of binary yes/no questions. The InCiSE model has undertaken its own analysis and aggregation of these measures to produce metrics for the crisis and risk management indicator. These are listed in detail in Annex A.

3.8.1 Imputation of missing data

One of the 38 countries selected for the 2019 edition of InCiSE has completely missing data for all crisis and risk management metrics. This is an improvement on the 2017 Pilot of InCiSE where eight countries had completely missing data. The 2017 Pilot used median imputation to handle missing data for imputing missing data for crisis and risk management. As a result of the decision to move to fully predictive imputation for the 2019 edition, external predictors have needed to be found. There are no easily identifiable external predictors (e.g. tertiary education for capabilities or the UN's E-Government survey for digital services), instead the correlations between the crisis and risk management metrics and other metrics in the InCiSE model have been analysed to identify potential predictors. This analysis has selected three metrics: the task discretion metric from the capabilities indicator; the use of data in HR administration from the HR management indicator; and, the Open Data Index from the openness indicator.

3.8.2 Changes from the 2017 Pilot

The 2017 Pilot used data solely from the national monitoring and progress reports of the UN Hyogo Framework for Action. The Hyogo Framework for Action ended in 2015 and has been replaced by the Sendai Framework, however monitoring and reporting of this framework has only just begun. Furthermore, these frameworks focus on natural disaster risk rather than the full range of risks and civil contingencies issues that countries have to manage at a central government level. Since the publication of the pilot a further dataset has become available, the OECD's Survey of the Governance of Critical Risks. This dataset provides data on this wider array of risks that governments, especially OECD members, tend to manage.

Table 3.8.A Composition of the crisis and risk indicator

Theme	Metric	Source	Type	Public sector proxy	Data transformation	Weighting within indicator		
						In theme (A)	Theme (B)	Total (c= A*B)
Strategic approach to risk	Approach (n)	OECD	Government assessment	No	Composite	20%	30%	6%
	Lead institution functions [new]	OECD	Government assessment	No	Composite	20%		6%
	Risk planning extent	UN	Government assessment	No	Composite	20%		6%
	Multi-hazard assessment	UN	Government assessment	No	Composite	20%		6%
	Risk monitoring	UN	Government assessment	No	Composite	20%		6%
Risk preparations and management	Risk management capability [new]	OECD	Government assessment	No	Composite	25%	30%	7.5%
	Preparedness	UN	Government assessment	No	Composite	25%		7.5%
	Disaster spending appraisal	UN	Government assessment	No	Composite	25%		7.5%
	International Cooperation	UN	Government assessment	No	Composite	25%		7.5%
Risk communications	Crisis and risk communications [new]	OECD	Government assessment	No	Composite	50%	20%	10%
	Early warning systems	UN	Government assessment	No	Composite	50%		10%
Post disaster assessment	Risk research and evaluation [new]	OECD	Government assessment	No	Composite	50%	20%	10%
	Post disaster assessment	UN	Government assessment	No	Composite	50%		10%

Table 3.8.B Definition of the crisis and risk metrics

Theme	Metric	Source	Definition of the source metric (e.g. question wording)
Strategic approach to risk	Approach (n)	OECD	Strategic level approaches to risk management
	Lead institution functions *[new]*	OECD	Functions carried out by the lead institution
	Risk planning extent	UN	Types of risk planning undertaken
	Multi-hazard assessment	UN	Approach to multi-hazard assessment
	Risk monitoring	UN	Extent of risk monitoring
Risk preparations and management	Risk management capability *[new]*	OECD	Approach to ensuring capability in crisis and risk management
	Preparedness	UN	Approach and policies for preparedness
	Disaster spending appraisal	UN	Approach to economic appraisal of disaster spending
	International Cooperation	UN	Processes and activities in place for international cooperation
Risk communications	Crisis and risk communications *[new]*	OECD	Approaches to communicating information about risks before and during events
	Early warning systems	UN	The existence and use of early warning systems
Post disaster assessment	Risk research and evaluation *[new]*	OECD	Use of research and evaluation in risk management policy and planning
	Post disaster assessment	UN	Approach to assessing post-disaster losses and needs

3.9 Procurement

The procurement indicator is defined as: the extent to which the government's procurement processes are efficient, competitive, fair and pursues value for money. According to the World Trade Organisation, "government procurement accounts for an average of 15 percent of more of a country's GDP" (2015). As procurement makes up such a large proportion of countries' GDP, it must be managed appropriately. Effective procurement management can streamline contracts and reduce outgoings, contributing to improved efficiencies in civil services. On public procurement, the World Bank (2016) states it "is a key variable in determining development outcomes and, when carried out in an efficient and transparent manner, it can play a strategic role in delivering more effective public services. It can also act as a powerful tool for development with profoundly positive repercussions for both good governance and more rapid and inclusive growth".

The procurement indicator is comprised of six metrics. This indicator is new for the 2019 edition of the index, and was not included in the 2017 Pilot edition. The sources for the procurement indicator are:

- The OECD's Public Procurement Survey [2016].

- Opentender (OT) analysis of European public procurement data by Digiwhist (a collaboration of the University of Cambridge, Open Knowledge Foundation Germany, Government Transparency Institute, Hertie School of Governance, Datlab and Transcrime) [2016].

3.9.1 Imputation of missing data

Two of the 38 countries selected for the 2019 edition of InCiSE have completely missing data for the procurement indicator. The procurement indicator is a new indicator for the 2019 edition, and there are no easily identifiable external predictors (e.g. tertiary education for Capabilities or the UN's E-Government survey for Digital Services), instead the correlations between the procurement metrics and the other metrics in the InCiSE model have been analysed to identify potential predictors. This analysis has selected three metrics: the use of data in HR administration from the HR management indicator; the publicised laws metric from the openness indicator; and, the collection cost metric from the tax administration indicator.

3.9.2 Changes from the 2017 Pilot

The procurement indicator is a new indicator and was not covered by the 2017 Pilot edition of the InCiSE Index.

Table 3.9.A Composition of the procurement indicator

Theme	Metric	Source	Type	Public sector proxy	Data transformation	Weighting within indicator		
						In theme (A)	Theme (B)	Total (c= A*B)
Procurement system	E-procurement functions (n)	OECD	Government assessment	No	Composite	33% (1/3)	50%	16.7% (1/6)
	Role of central purchasing body (n)	OECD	Government assessment	No	Composite	33% (1/3)		16.7% (1/6)
	Access for SMEs (n)	OECD	Government assessment	No	Composite	33% (1/3)		16.7% (1/6)
Procurement in practice	Admin effectiveness (n)	OT	Analysis of published data	Yes	None	33% (1/3)	50%	16.7% (1/6)
	Transparency (n)	OT	Analysis of published data	Yes	None	33% (1/3)		16.7% (1/6)
	Integrity (n)	OT	Analysis of published data	Yes	None	33% (1/3)		16.7% (1/6)

Table 3.9.B Definition of the procurement metrics

Theme	Metric	Source	Definition of the source metric (e.g. question wording)
Procurement system	E-procurement functions [new]	OECD	Are the following functionalities provided by the e-procurement system?
	Role of central purchasing body [new]	OECD	Please identify the roles of central purchasing bodies in your country
	Access for SMEs [new]	OECD	Please identify the approaches in place in your country to support participation in small and medium-sized enterprises (SMEs) in public procurement for central government.
Procurement in practice	Admin effectiveness [new]	OT	Made up of indicators: joint or centralised procurement, use of framework agreements, use of WTO framework, English as foreign language, Discrepancies between call for tender and contract award notices.
	Transparency [new]	OT	Share of non-missing key fields from tenders.
	Integrity [new]	OT	Made up of indicators: single bidder contract, call for tenders publication, procedure type, length of advertisement period, length of decision period, new company, tax haven

3.10 HR management

The HR Management indicator is defined as: the meritocracy of recruitment and extent to which civil servants are effectively attracted, managed and developed. "The public sector is very labour intensive – around 70 per cent of the budgets of most public organisations are spent on staff" (Bovaird and Löffler, 2003), so good HR management is key to the successful functioning of an exemplary civil service. Performance management can help create incentives for personal development in the civil service. Fukuyama (2013) recognises that recruitment and reward "remain at the core of any measure of quality of governance. Whether bureaucrats are recruited and promoted on the basis of merit". Meanwhile, Bovaird and Löffler (2003) note that "if the HR policies are not right, then public organisations will not attract the human resources they need to perform the functions of government and deliver the services that government has promised the electorate".

The HR management indicator is comprised of nine metrics, an increase of four from the 2017 Pilot. The data sources for the indicator are:

- Quality of Government expert survey by the University of Gothenburg (QoG) [2015].

- OECD survey on Strategic Human Resources Management (OECD) [2016].

3.10.1 Imputation of missing data

None of the 38 countries selected for the 2019 edition of InCiSE have completely missing data for the HR management metrics. As a result the imputation of missing data for the HR management metrics is based solely on the data within the indicator.

3.10.2 Changes from the 2017 Pilot

In the 2017 Pilot, InCiSE used five metrics from the Quality of Governance study. These provided only partial coverage of the topic area, with a particularly strong focus on meritocratic recruitment. Since the 2017 Pilot, the OECD published the 2017 edition of their Government at a Glance report, including a number of measures from their 2016 Survey on Strategic Human Resource Management. The 2019 edition of InCiSE has incorporated three metrics from this survey as published in Government at a Glance in order to improve the coverage of the indicator.

While there continue to be arguments about the use and implementation of performance appraisal and performance-related pay mechanisms within public sector organisations, the OECD (2005) suggests that even if there is no direct performance improvement associated with these measures they can act as a catalyst for change. Thus, there may be secondary effects from performance appraisal and performance related pay that improve civil service effectiveness.

Table 3.10.A Composition of the HR management indicator

Theme	Metric	Source	Type	Public sector proxy	Data transformation	In theme (A)	Theme (B)	Total (c= A*B)
Meritocratic recruitment	Applicant skills	QoG	Expert assessment	Yes	None	25%	50%	12.5%
	Political connection	QoG	Expert assessment	Yes	Inverted	25%		12.5%
	Personal connection	QoG	Expert assessment	Yes	Inverted	25%		12.5%
	Formal recruitment	QoG	Expert assessment	Yes	None	25%		12.5%
Attracting and retaining talent	Comparable salaries	QoG	Expert assessment	Yes	None	50%	20%	10%
	Performance related pay [new]	OECD	Government assessment	No	None	50%		10%
HR practices	Use of performance assessment [new]	OECD	Government assessment	No	None	50%	20%	10%
	SCS HRM practices [new]	OECD	Government assessment	No	None	50%		10%
Data-driven HR	HR databases [new]	OECD	Government assessment	No	None	100%	10%	10%

*(Weighting within indicator spans the "In theme (A)", "Theme (B)" and "Total (c= A*B)" columns.)*

Table 3.10.B Definition of the HR management metrics

Theme	Metric	Source	Definition of the source metric (e.g. question wording)
Meritocratic recruitment	Applicant skills	QoG	When recruiting public sector employees, the skills and merits of the applicants decide who gets the job? [1=hardly ever, 7=almost always; mean score]
	Political connection	QoG	When recruiting public sector employees, the political connections of the applicants decide who gets the job? [Rated 1-7; mean score]
	Personal connection	QoG	When recruiting public sector employees, the personal connections of the applicants (for example kinship or friendship) decide who gets the job? [Rated 1-7; mean score]
	Formal recruitment	QoG	Public sector employees are hired via a formal examination system? [Rated 1-7; mean score]
Attracting and retaining talent	Comparable salaries	QoG	Senior officials have salaries that are comparable with the salaries of private sector managers with roughly similar training and responsibilities? [Rated 1-7; mean score]
	Performance related pay [new]	OECD	Index of the use of performance-related pay [Index ranging from 0 to 1]
HR practices	Use of performance assessment [new]	OECD	Index of the use of performance assessment in HR decisions [Index ranging from 0 to 1]
	SCS HRM practices [new]	OECD	Index of whether distinct HR practices are used for senior civil servants [Index ranging from 0 to 1]
Data-driven HR	HR databases [new]	OECD	Index of the use of databases for HR management [Index ranging from 0 to 1]

3.11 Tax administration

The tax administration indicator is defined as: the efficiency and effectiveness of tax collection (at the central/federal level). Effective tax systems can be viewed as a critical building block for increased domestic resource mobilisation which is essential for civil service effectiveness and good governance. "Successful tax extraction provides resources that enable the government to operate in other domains", Fukuyama (2013) highlights "it is a necessary function of all states, and one for which considerable data exist". The role of tax administration as the basis of government operations is made clear by the OECD (n.d): "Strong tax administrations and sound public financial management help maximise the domestic resources that are necessary for government to function, to sustain social safety nets, to maintain long-term fiscal sustainability, and to free up fiscal space for pursuing socio-economics objectives". Although priorities and circumstances vary widely across countries, the drive to elevate the collective standard of tax administration is of great importance. Holt and Manning highlight the importance of tax administration in measuring the effectiveness of public administration and it is one of the key functions highlighted by the World Bank Indicators of the Strength of Public Management Systems (2012).

The tax administration indicator is comprised of six metrics and its structure is unchanged from the 2017 Pilot edition of InCiSE. The data sources for the indicator are:

- OECD's Tax Administration Comparative Information Series [2015].

- The World Bank's 'Doing Business' Index (WB) [2018].

3.11.1 Imputation of missing data

None of the 38 countries selected for the 2019 edition of InCiSE have completely missing data for the tax administration metrics. As a result the imputation of missing data for the tax administration metrics is based solely on the data within the indicator.

3.11.2 Changes from the 2017 Pilot

There are no changes to the structure of the tax administration indicator.

Table 3.11.A Composition of the tax administration indicator

Theme	Metric	Source	Type	Public sector proxy	Data transformation	Weighting within indicator		Total (c= A*B)
						In theme (A)	Theme (B)	
Efficiency of collection	Collection Cost	OECD	Government assessment	No	Composite	50%	33%	16.67% (1/6)
	Tax Debt	OECD	Government assessment	No	Composite	50%		16.67% (1/6)
Taxpayer experience	Time to Pay Taxes – Business	WB	Expert assessment	No	Inverted	100%	33%	33%
Digital tax services	E-Filed Tax Returns – Personal	OECD	Government assessment	No	Composite	33%	33%	11.11% (1/9)
	E-Filed Tax Returns – Corporate	OECD	Government assessment	No	Composite	33%		11.11% (1/9)
	E-Filed Tax Returns – VAT	OECD	Government assessment	No	Composite	33%		11.11% (1/9)

Table 3.11.B Definition of the tax administration metrics

Theme	Metric	Source	Definition of the source metric (e.g. question wording)
Efficiency of collection	Collection Cost	OECD	Administrative costs as a proportion of net revenue
	Tax Debt	OECD	Total tax debt as a proportion of net revenue
Taxpayer experience	Time to Pay Taxes – Business	WB	The time it takes businesses to pay taxes
Digital tax services	E-Filed Tax Returns – Personal	OECD	The percentage of tax returns e-filed during the last fiscal year for personal taxes
	E-Filed Tax Returns – Personal	OECD	The percentage of tax returns e-filed during the last fiscal year for corporate taxes
	E-Filed Tax Returns – VAT	OECD	The percentage of tax returns e-filed during the last fiscal year for VAT taxes

3.12 Digital services

The digital services indicator in InCiSE is defined as the user-centricity and cross-border mobility of digitally-provided public services and the availability of 'key enablers'. A changing world and digital environment provide the impetus for a civil service to ensure modernity and remain user-centric for the public. In doing so, efficiencies should be achieved to enable cost savings in processes while also allowing for further accessibility of services. The OECD has supported this view of potential benefits: "ICT is increasingly used to support broader public sector development objectives… by changing service delivery approaches by creating personalised, high quality services to users, thereby increasing user satisfaction and effective service delivery; facilitating major work organisation and management changes creating back-office coherence and efficiency gains; increasing transparency of government activities, and increasing citizen engagement" (Lontii and Woods, 2008).

The source data for the digital services indicator is the European Commission's eGovernment Benchmark Report (eGBR) 2017 and 2018 reports, which provide data for 2016 and 2017 respectively. This is the same source that was used in the 2017 Pilot, however significant changes have been made to the way in which the data is extracted and imported. The 2019 edition of the digital services indicator is composed of 13 metrics, compared to four in the 2017 Pilot.

3.12.1 Imputation of missing data

Nine of the 38 countries selected for the 2019 edition of InCiSE have completely missing data for the digital services indicator. The 2017 Pilot of the InCiSE Index set out the use of Online Services Index from the UN's biennial E-Government Survey as the external predictor for imputation. This approach is maintained for the 2019 edition of the InCiSE Index.

3.12.2 Changes from the 2017 Pilot

The data source used for the digital services indicator in the 2019 edition of InCiSE is the same as that used for the 2017 Pilot – the European Commission's eGovernment Benchmark Report (eGBR). However, further investigation of the data and methodology of the report has led to a change in the metrics used by InCiSE. While the 2017 Pilot took four high-level metrics, the 2019 edition of InCiSE will use 13 more granular metrics.

The eGBR uses mystery shopping of eight 'life events' to assess the quality of digital public services in all 28 EU member countries and six other neighbouring/partner countries. These life events are designed to capture the majority of interactions that citizens and businesses have with public services in European nations. The services assessed by the eGBR include not only national level services but also those provided by subnational and local governments. As InCiSE aims to look at the effectiveness of national-level civil services we investigated whether there was a way to exclude non-national services.

While the European Commission publishes the full underlying data for the eGBR, it is not easy to calculate scores based solely on the assessments of national-level services. So, an analysis of the data from the 2016 and 2017 reports was undertaken to look at the pattern of service delivery across the eight life events. The results of this analysis is presented in Table 3.12.C, and shows that for five of the eight life events more than half of the URLs assessed by the eGBR are recorded as 'national' level services. However, for the 'moving house', 'owning and driving a car' and 'studying' life events the analysis shows that in most countries the URLs being assessed are sub-national/local services.

Table 3.12.A Composition of the digital services indicator

Theme	Metric	Source	Type	Public sector proxy	Data transformation	Weighting within indicator		Total (c= A*B)
						In theme (A)	Theme (B)	
User experience	Business: start-up *[new]*	eGBR	Expert assessment	Yes	None	20%	40%	8%
	Business: regular operations *[new]*	eGBR	Expert assessment	Yes	None	20%		8%
	Family life *[new]*	eGBR	Expert assessment	Yes	None	20%		8%
	Finding and losing a job *[new]*	eGBR	Expert assessment	Yes	None	20%		8%
	Small claims procedure *[new]*	eGBR	Expert assessment	Yes	None	20%		8%
Key enablers	Business: start-up *[new]*	eGBR	Expert assessment	Yes	None	20%	40%	8%
	Business: regular operations *[new]*	eGBR	Expert assessment	Yes	None	20%		8%
	Family life *[new]*	eGBR	Expert assessment	Yes	None	20%		8%
	Finding and losing a job *[new]*	eGBR	Expert assessment	Yes	None	20%		8%
	Small claims procedure *[new]*	eGBR	Expert assessment	Yes	None	20%		8%
Cross-border services	Business: start-up *[new]*	eGBR	Expert assessment	Yes	None	33% (1/3)	20%	6.67% (1/15)
	Business: regular operations *[new]*	eGBR	Expert assessment	Yes	None	33% (1/3)		6.67% (1/15)
	Small claims procedure *[new]*	eGBR	Expert assessment	Yes	None	33% (1/3)		6.67% (1/15)

Table 3.12.B Definition of the digital services metrics

Theme	Metric	Source	Definition of the source metric (e.g. question wording)
User experience	Business: start-up [new]	eGBR	Online availability and usability of services to start a business and initial operations
	Business: regular operations [new]	eGBR	Online availability and usability of services for regular business operations
	Family life [new]	eGBR	Online availability and usability of services for registering births, marriages and deaths
	Finding and losing a job [new]	eGBR	Online availability and usability of services to help when you lose a job and for finding a new job
	Small claims procedure [new]	eGBR	Online availability and usability of services for initiating small claims/magistrates procedures
Key enablers	Business: start-up [new]	eGBR	Use of authentication and other enabling functions for business start-up and initial operations
	Business: regular operations [new]	eGBR	Use of authentication and other enabling functions for regular business operations
	Family life [new]	eGBR	Use of authentication and other enabling functions for registering births, marriages and deaths
	Finding and losing a job [new]	eGBR	Use of authentication and other enabling functions for services when lost/finding a job
	Small claims procedure [new]	eGBR	Use of authentication and other enabling services for small claims/magistrates procedures
Cross-border services	Business: start-up [new]	eGBR	Ability for foreigners to use business start-up services
	Business: regular operations [new]	eGBR	Ability for foreigners to undertake regular business operations
	Small claims procedure [new]	eGBR	Ability for foreigners to use small claims/magistrate procedures

For each of the eight life events the mystery shopping exercise looks across three domains: 'user centric government', 'transparency' and 'key enablers'; six of the eight life events are also assessed for the additional domain of 'cross-border mobility'. As transparency is already covered in InCiSE through the openness indicator, including the eGBR transparency data could be seen as duplicating information already measured elsewhere in the InCiSE framework.

Therefore, in the 2019 edition of InCiSE rather than use the high-level averages for the four domains (as used in the 2017 Pilot), the model uses the 'user centric', 'transparency' and 'key enablers' domain scores for the business start-up, regular business operations, family life, losing and finding a job, and small claims procedure life events. This approach removes scores for the three life events (moving house, owning and driving a car, and studying) where services are typically not delivered by national governments, and reduces potential overlap with the openness indicator by removing scores for the 'transparency' domain.

Table 3.12.C Proportion of eGBR assessed services identified as 'national' level services

Life event	Median proportion of assessed URLs that are for 'national' services	Countries where less than 50% of assessed URLs are for national services (out of 34)
Business: start-up and early trading	91%	3
Regular business operations	83%	4
Family life	61%	13
Losing and finding a job	86%	4
Moving house	23%	28
Owning and driving a car	49%	17
Small claims procedures	66%	11
Studying	37%	25

Chapter 4: Summary of changes from the 2017 Pilot edition of the InCiSE Index

The 2019 edition of InCiSE incorporates a number of methodological changes and improvement since the 2017 Pilot, which are the result of desk research, stakeholder feedback and engagement since the pilot publication. This chapter provides a general summary of the changes since the 2017 Pilot.

4.1 Changes in the overarching methodology

There are two main changes to the overarching methodological approach for the 2019 edition of InCiSE. Firstly, the technical modelling is being done in the R software package, rather than the mix of Excel and Stata that was used for the pilot. This approach reduces the potential for error, while the use of open source software will increase the opportunities for reproducibility. Secondly, a 'data quality assessment' has been introduced which makes a quantitative appraisal of the data quality of countries and indicators. This assessment has been used to determine country selection, and to partially account for data quality in the weighting of the indicators into the composite index score.

4.2 Indicators with no changes

There are three indicators with no changes to their definition or metrics – policy making, inclusiveness and tax administration. For policy making and tax administration there have been data updates to all metrics, while two of the five metrics in inclusiveness have been updated.

4.3 Indicators with minor changes

There are five indicators with what we class as 'minor' changes, that is changes that we do not believe substantially change or which are not contentious. For the openness, integrity and regulation indicators we have identified some additional metrics in the Bertelsmann Foundation's Sustainable Governance Indicators that enhance the topic coverage of these indicators. For the integrity indicator we are also making a change to the coding of post-employment cooling-off periods to remove consideration of whether compensation is paid during the cooling-off period due to quality concerns about this aspect of the data. For the fiscal & financial management indicator we are adding three metrics (one from the International Budget Partnership and two from the World Bank) that measure government's openness/publication of budget and public spending documents and statistics. For the HR management indicator we are incorporating newly published data from the OECD on strategic HRM practices.

4.4 Crisis and risk management

The crisis and risk management indicator has been redesigned, drawing from both the 2017 Pilot source (the Hyogo Framework for Action monitoring reports) and new data from the OECD on the governance of critical risks. The 2017 Pilot data focuses heavily on natural disaster risk management, the OECD data substantially enhances the topic coverage and provide a more rounded view of crisis and risk management practices.

4.5 Capabilities

A data quality concern about the capabilities indicators is that the data for most countries has a reference date of 2012. It has not been possible to identify new and more up-to-date data for the capabilities indicator (the source data is the OECD Survey of Adult Skills), although further datasets for this data source that expand country coverage for this indicator were identified. This led to a further review of the source data, which led to the identification of a range of additional metrics that could be incorporated into the model. The metrics in the pilot focused on capability levels (literacy, numeracy, problem solving skills, and education level), however the data also includes a number of metrics on the use of skills and learning at work (e.g. use of reading/writing/IT skills at work, formal and informal learning for job-related reasons in the past 12 months). Furthermore, the pilot used data for the public sector as a whole, however investigation of the source data suggested that reliable estimates for the 'public administration' industrial sector could be produced (this is wider than just the civil service, including things like local government, but excluding things such as healthcare, education and transport). The capabilities indicator has therefore incorporated 10 additional metrics on skills use and learning at work, and switched to using data for the 'public administration' industrial sector.

4.6 Digital services

The source data for digital services (the European Commission's eGovernment Benchmark Report) uses a 'life events' model, however for a number of these life events delivery across the countries included in the dataset is at the sub-national/local level. Moreover, one of the domains (transparency) overlaps with an existing InCiSE indicator. Therefore, the way in which data is extracted has been changed to select data for those life events where for a majority of countries the service is delivered at the national level (and therefore likely to be managed by the civil service) and to exclude the transparency domain.

4.7 Procurement

Since the 2017 Pilot, two data sources have been identified that can provide metrics for an indicator on procurement (an element of the InCiSE framework not covered by the pilot). One source is the OECD's Survey on Public Procurement which looks at the role of CPBs and strategic approaches to public procurement (e.g. e-procurement and support for SMEs). The other source is the Opentender project, supported by an academic consortium, which analyses the tender and contract notices for procurement exercises using the European Union's Tenders Electronic Daily service.

4.8 Social security

The 2017 Pilot included an indicator for social security. This was based on a single metric: administrative costs as a proportion of total social protection spending. Feedback received following the publication of the pilot identified significant quality issues with the metric used. No alternative metrics for the indicator were identified, therefore it was decided to depreciate the indicator from the model. Further discussion of this is provided in Chapter 6.

Chapter 5: Sensitivity analysis

Building statistical models and indices involves stages where subjective judgements have to be made. These can include the selection of individual data sets, the treatment of missing values, and the approach to weighting and aggregation. Good modelling practice means we should evaluate our model, testing the assumptions and judgements made in its building and analysing the uncertainties associated with the modelling process. Sensitivity analysis is one way to undertake such an assessment.

To test the robustness and uncertainty of the modelling approach used by InCiSE, five types of sensitivity analysis have been undertaken:

- Varying the set of countries selected for results to be produced;

- Excluding out-of-date data;

- Alternative approaches to weighting;

- Using the ranks of source data; and,

- Alternative approaches to imputation.

This chapter summarises the approach and results of these different analyses, while detailed results can be found in Annex B.

5.1 Country selection
Section 2.3 discusses how the approach to country selection for the 2019 edition of InCiSE differs from the 2017 Pilot, as

it now uses the results of the data quality assessment (DQA) to identify countries for inclusion. The DQA produces a score for each country that summarises the quality of the data within the InCiSE model about that country (before imputation of missing values). The threshold for inclusion in the 2019 edition of InCiSE is an overall DQA score of 0.50 or greater.

The three countries included in the InCiSE Index with the lowest data quality scores have markedly poorer data quality by indicator than other countries (see Table 2.8.A). For each of these three countries only two or three of the 12 InCiSE indicators are rated green, a further two or three indicators are rated as amber, while five or six are rated as red, and one indicator is fully imputed. Section 2.8 also outlines an approach to 'grading' countries based on their data quality scores. DQA scores of 0.75 are given an 'A+' grade, while those below 0.6 are given a 'D' grade. In this 'D' group there are four more countries in addition to the three discussed above.

The 2017 Pilot used a simpler approach to country inclusion with a threshold of having at least 75% of metrics available, and producing a set of 31 countries.[5] For the 2019 edition's set of metrics 31 countries also achieve the 75% threshold but the country coverage differs to the set of countries in the 2017 Pilot.

---

5   One further country in 2017 met this criteria but was not an OECD member so was excluded to simplify interpretation of results.

The first two sensitivity tests for country coverage altered the DQA threshold used to determine country inclusion. The first test used a DQA score of 0.55 or higher, excluding the three countries in the 2019 set with the lowest data quality, while the second test used a DQA score of 0.6 or higher. The third test used the 2017 Pilot's threshold of countries with 75% of data being available. The fourth test used the 31 countries included in the 2017 Pilot.

5.2 Reference date

The reference dates of the source data for the 2019 edition of InCiSE ranges from 2012 to 2018. However, as shown in Table 5.2.A, the reference dates vary across indicators. A third of the metrics have a reference date of 2017 or 2018, around half of the metrics

have a reference date of 2015 or 2016, while just 17 out of the 116 metrics have a reference date of 2012.

Of these 17 metrics, 14 are the metrics for the capabilities indicator. This is the only indicator with 100% of its data with a reference date from before 2015.[6] The capabilities indicator is solely composed of data with a reference year of 2012. Only two other indicators have data from before 2014 but in both cases this is a small number of their constituent metrics.

The first two sensitivity tests for recency exclude the capabilities indicator. In the first analysis the capabilities indicator is excluded but the weightings of the other indicators are not adjusted. In the second analysis the

---

6   The lack of recency of the data source for the capabilities indicator (the OECD's Survey of Adult Skills) is discussed in section 3.3.

Table 5.2.A Reference year of InCiSE metrics by indicator

	Number of metrics per year							Percent within in period...		
	2012	2013	2014	2015	2016	2017	2018	2012-14	2015-16	2017-18
Capabilities	14							100%		
Crisis & risk management				8	5				100%	
Digital services					7	6			54%	46%
Fiscal & financial management	1				1	4		17%	17%	67%
HR management				5	4				100%	
Inclusiveness				3	2				100%	
Integrity		1	2	11		2	1	18%	65%	18%
Openness				1	3	4	2		40%	60%
Policy making						8				100%
Procurement					6				100%	
Regulation						6	3			100%
Tax administration				5		1			83%	32%
Total	15	1	2	33	28	23	14	16%	52%	32%

weightings are recalculated to account for the removal of the capabilities indicator.

In the third test, only data with a reference year of 2015 or later is included in the model; the four other metrics from before 2014 are excluded in addition to the 14 capabilities metrics. In the fourth test, only data with a reference year of 2016 or later is included in the model; the 51 metrics with a reference date of 2016 or earlier are therefore excluded. For both these analyses there is no adjustment the weightings – either to calculate the indicators from their constituent metrics or to calculate the index from the indicators.

5.3 Alternative approaches to weighting

The InCiSE Index is a weighted aggregation of the InCiSE indicators, which themselves are weighted aggregations of the InCiSE metrics. Section 2.7 set out the approach to weighting the InCiSE indicators to calculate the InCiSE Index. Two-thirds of an indicator's weight is based on an 'equal share' approach (i.e. 1/12), while one-third is based on the results of the data quality assessment. Section 2.6 and Chapter 3 outline how the metrics are weighted to produce each of the 12 indicator scores.

The first three sensitivity tests for alternative weighting look at the proportion of indicator weighting that is assigned to the 'equal share' and the data quality assessment. The first test uses a 50:50 split rather than the 67:33 split. The second test uses solely an 'equal share' approach (i.e. indicator weights set to 1/12 each). The third test uses solely the results of the data quality assessment to determine the weighting.

The fourth and fifth tests focus on metrics weighting: The fourth does not apply weighting to metrics within indicators (i.e.

all metrics contribute equally to the calculation of their indicator), and the fifth is a simple summation of the metrics, then normalised as per the standard calculations of the indicators and index (as set out in section 2.5).

5.4 Adjusting the base data
In the InCiSE model, metrics are normalised after missing data is imputed. An alternative approach would be to normalise the data before it is imputed.

Three sensitivity tests were done where normalisation of the data occurred before the imputation. In the first test the data was ranked, in the second test the data was rescaled using the same min-max normalisation applied to the outputs of the model, and in the third test the data was converted to z-scores with a mean of 0 and a standard deviation of 1.

5.5 Alternative imputation methods
As discussed in section 2.4 missing data in the InCiSE base data is handled through multiple imputation, and in particular the predictive mean matching method.

Four sensitivity tests were carried out using different approaches to imputation. Section 2.4 outlines how the imputation of missing data is handled on a per-indicator basis, the first test changes this to adopt a "kitchen sink"/"all-in-one" approach in which the full dataset of all 116 metrics (and two external predictor variables) are supplied to the imputation function. The second test uses a modified form of predictive mean matching called 'midas touch' to generate imputed values. The third test uses the 'random forest' method to generate imputed values, a machine learning approach. The fourth test uses mean imputation, where missing data is replaced with the simple arithmetic mean of the observed data.

5.6 Results of the sensitivity analysis

Table 5.6.A shows the results of the 2019 InCiSE model for each country and the range of ranks across the five different sets of sensitivity analysis, while Figure 5.1 to 5.5 show how the InCiSE Index score varies by country for each of the sensitivity tests carried out. The results of the five sets of sensitivity analysis demonstrate general stability in the model, with country ranks either unchanged or changed by only one or two places on average, and the same groupings of countries at the top and bottom of the rankings. Full results from the sensitivity analysis are provided in Annex B.

In the country coverage sensitivity analysis, the main driver of change in rankings is due to the exclusion of countries: Figure 5.1 shows that the scores of individual countries do not substantially change as a result of the exclusion of different countries. When varying the reference date there are some changes as a result of the exclusion of the capabilities indicator, and further changes as a result of excluding data with a reference year of 2015 and earlier.

Altering the weighting schemes for the calculation of the index and indicators does not result in many changes, except when calculating the index as a simple sum of all metrics (i.e. applying no weighting at all). Similarly making alterations to the metrics (e.g. ranking, rescaling, standardisation) before they are imputed does not result in many changes to country scores or rankings.

Varying the imputation methodology results in slightly more variation of country scores and ranks than the previous sensitivity checks. Only three countries see no change in their ranking, however of those that do change, the difference in ranks is still small at around one or two places.

One way to consider the effectiveness of the sensitivity analysis is to calculate the Mean Absolute Error (MAE) arising from the analysis. MAE is a common technique for assessing the quality of statistical models by comparing the difference of the model's estimates/predictions with the original data. It is calculated as the sum of the absolute errors divided by the number of cases. In the case of the InCiSE sensitivity analysis, 'error' is calculated as the difference between the 2019 InCiSE Index results and the results from each of the sensitivity tests.

The overall MAE figure for the sensitivity analysis, that is the mean level of 'error' across all 20 sensitivity tests for all 38 countries, is ±0.017.
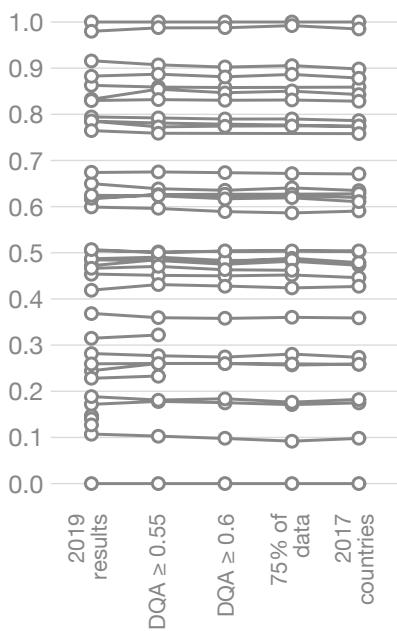
The MAE can also be calculated for each sensitivity test or each set of tests. The per-set MAE figures is presented at the bottom of Table 5.6.A, while the per-test MAE is presented in the tables in Annex B. Across the different sets of methodological sensitivity tests, the smallest MAE is ±0.007 for the set of tests varying country selection while the highest MAE is ±0.023 for the set of tests changing the reference date.

Finally, the MAE can also be calculated by country, which is also included in Table 5.6.A and ranges from ±0.001 to ±0.032. However, given that the same two countries place highest and lowest across most tests the minimum per-country MAE is skewed by the limited variability in these two countries' scores, when excluding these countries the minimum MAE rises from ±0.001 to ±0.009.
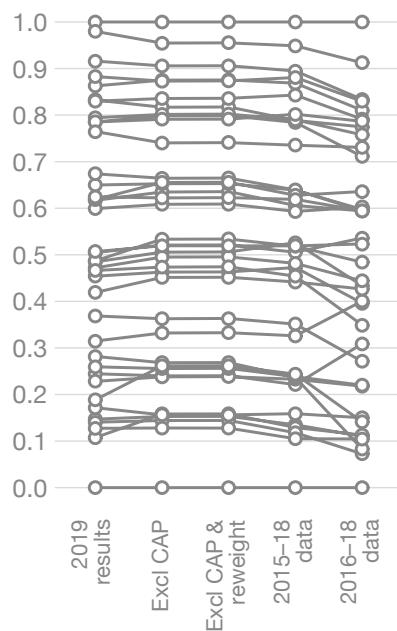
Table 5.6.A Variation in country ranking across sensitivity analyses

Country	2019 results		Range of country's rank in sensitivity analysis					Mean Absolute Error
	Score	Rank	Country coverage	Reference date	Alternative weightings	Adjusting base data	Imputation method	
GBR	1.000	1	1	1	1-2	1	1-2	0.003
NZL	0.980	2	2	2	1-2	2	1-2	0.019
CAN	0.916	3	3	3	3	3	3-5	0.021
FIN	0.883	4	4	4-5	4-5	4	3-4	0.013
AUS	0.863	5	5	4-5	4-5	5-6	4-7	0.014
DNK	0.832	6	5-6	7-9	6-8	5-7	5-7	0.021
NOR	0.830	7	6-7	6	6-7	6-10	5-7	0.010
NLD	0.794	8	7-8	8-9	8-10	8-9	8-9	0.014
KOR	0.785	9	8-10	9-11	6-11	7-11	10	0.019
SWE	0.785	10	9-10	7-10	8-10	8-9	8-9	0.009
USA	0.765	11	11	10-11	10-11	10-11	11	0.028
EST	0.674	12	10-12	12-17	12	12-13	12-15	0.023
CHE	0.650	13	11-13	13-14	13-14	12-15	12-15	0.020
IRL	0.625	14	14-16	15-16	14-17	14-15	16-17	0.021
FRA	0.619	15	12-15	12-14	13-16	13-15	12-15	0.012
AUT	0.617	16	13-15	15-16	13-16	16-17	13-15	0.014
ESP	0.599	17	15-17	13-17	15-17	16-17	16-17	0.010
MEX	0.507	18	17-19	19-20	18-24	18-23	18-20	0.020
DEU	0.505	19	16-19	18-21	18-19	19-21	18-20	0.010
LTU	0.487	20	18-20	18-20	20-22	20-21	20-22	0.018
BEL	0.485	21	19-22	18-22	20-21	19-20	18-21	0.017
JPN	0.472	22	17-21	21-22	19-24	18-23	21-24	0.020
LVA	0.466	23	20-23	23-26	20-24	24	24-26	0.032
CHL	0.454	24	21-24	23-25	22-24	22-23	21-23	0.014
ITA	0.419	25	22-25	23-25	25-26	25	23-25	0.014
SVN	0.369	26	23-26	26-28	25-26	26	25-26	0.018
ISR	0.315	27	27	24-27	27	27	27-29	0.022
POL	0.282	28	24-28	28-36	28-29	28-29	27-29	0.025
PRT	0.259	29	25-29	29-30	28-29	31	28-31	0.015
CZE	0.245	30	26-30	27-32	30-32	28-30	30-31	0.018
ISL	0.228	31	31	30-32	30-32	29-30	28-31	0.019
TUR	0.189	32	27-32	28-32	30-35	32	32-33	0.026
SVK	0.172	33	28-33	31-34	32-35	33	32-34	0.015
BGR	0.147	34	n/a	34-35	33-34	35	35-36	0.016
HRV	0.140	35	n/a	36-37	34-36	34	33-34	0.019
ROU	0.127	36	n/a	35-37	36-37	36	35-37	0.022
GRC	0.107	37	29-34	33-35	34-38	37	36-37	0.027
HUN	0.000	38	30-35	38	37-38	38	38	0.001
Mean Absolute Error (MAE)			0.007	0.023	0.014	0.014	0.022	
No change in rank			8	5	3	16	3	
Largest difference in rank			± 8	± 8	± 6	± 5	± 3	
Average difference in rank			± 2	± 2	± 2	± 1	± 2	
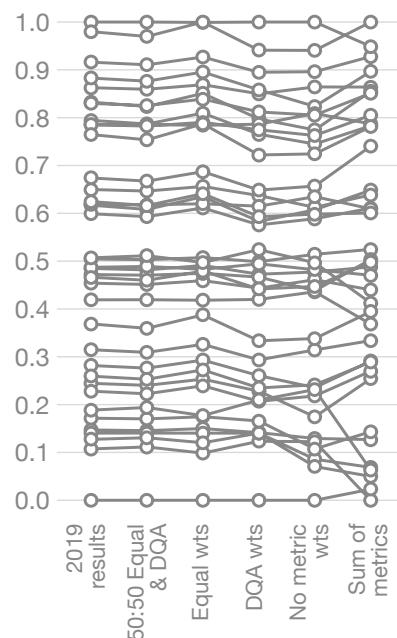
Figure 5.1
InCiSE Index scores using different country groupings
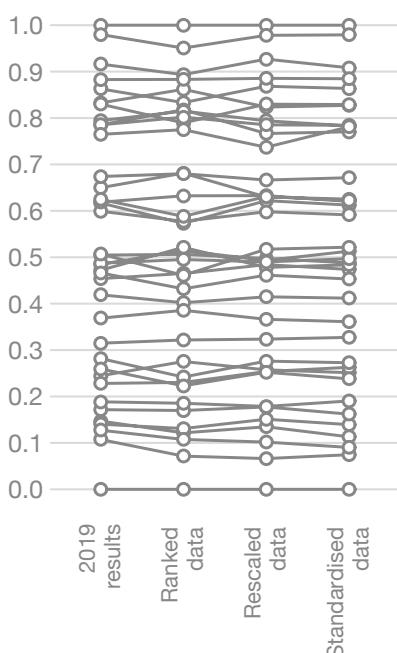
Figure 5.2
InCiSE Index scores excluding 'out of date' data
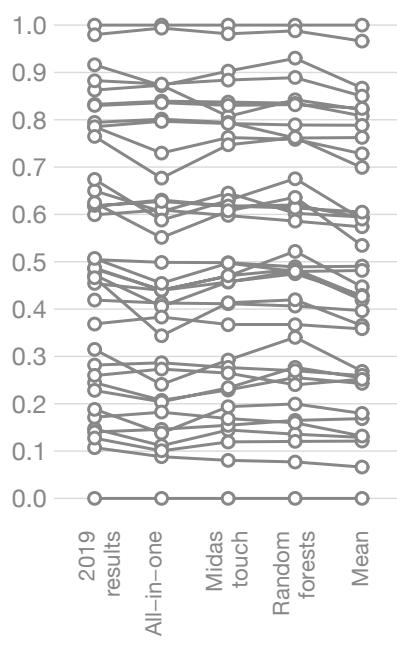
Figure 5.3
InCiSE Index scores using alternative weighting

Figure 5.4
InCiSE Index scores adjusting the base data

Figure 5.5
InCiSE Index scores using alternative imputation methods

Chapter 6: Future development

The 2019 index is the second edition of the InCiSE project, following the pilot edition published in 2017. The 2019 edition builds on and strengthens the methodology of the pilot edition. The InCiSE Partners have used a combination of stakeholder feedback, continued engagement with data providers and further desk research to develop the methodology for the 2019 edition of the InCiSE Index.

Given the frequency of data updates and to provide suitable time to reflect on each edition's results, we propose that future editions of the InCiSE Index are repeated on a biennial timescale. This chapter sets out considerations for future development of the InCiSE methodology.

6.1 Social security administration
The InCiSE framework (described in section 1.3) identifies social security administration as one of the constituent functions of an effective central civil service, and the 2017 Pilot edition of the InCiSE Index included an indicator for social security administration. The indicator was based on a single metric, which was the administrative costs of social protection as a proportion of total social protection expenditure. Feedback from the pilot edition included a critique of this metric, saying it was unsuitable given the inclusion of state provided healthcare which varies significantly across countries. Furthermore, the data was available solely for European Union member states, so data for non-EU countries was imputed based on correlated

perception measures from the Quality of Governance study used elsewhere in the InCiSE model.

Exploration of the source data did not identify an appropriate method to exclude healthcare costs from the calculations. A review of further data sources identified neither alternative metrics that included non-EU countries nor imputation predictors with a closer intellectual or theoretical relationship to the indicator's conceptual basis.

It was therefore decided that the social security indicator should be removed from the 2019 edition of InCISE. For future editions of the InCiSE Index, we will continue to explore whether there is suitable data to reintroduce a social security indicator.

6.2 Functions and attributes not yet measured
In addition to social security, four of the functions and attributes identified in the InCiSE framework have not been measured in either edition of the index: IT for officials, internal finance, staff engagement, and innovation. No suitable data has been identified since the pilot that would allow for measurement of these four potential indicators. Future editions of the InCiSE index will continue to explore whether suitable data exists to introduce indicators for these four areas.

6.3 Functions and attributes already measured

The 2019 edition of InCiSE has used an additional 46 metrics compared to the 2017 Pilot: six form the new procurement indicator and 40 are distributed across the existing indicators measured in the 2017 Pilot. While this has strengthened a number of indicators, as Table 2.8.B shows only three of the indicators have been given a final 'RAG' rating of green (data quality score of 0.75 or more). Table 6.1 below provides some considerations for future improvements of each of the indicators measured in the 2019 edition of InCiSE, with amber or red 'RAG' ratings.

6.4 Extending country coverage

While coverage of the InCiSE results has increased from the 31 countries in the 2017 Pilot to 38 in the 2019 edition, the group of countries remains broadly homogenous, made up of OECD and EU member countries with high or upper-middle incomes. Future editions of the InCiSE Index will continue to use the data quality based approach

to country inclusion set out in section 2.3, however this requires greater data availability for non-OECD/EU countries.

There are a number of potential options, such as creating regional versions of the InCiSE Index using existing multi-country data collections for different regions (but for which either OECD or EU countries are not members). Alternatively, subsets of the existing InCiSE Index could be created as some indicators have wider data coverage than others.

The InCiSE Partners are committed to identifying ways to increase coverage, and have conducted two short studies of how the InCiSE framework applies in Brazil and Nigeria to inform future thinking.

While extending country coverage will generate a greater set of results, careful consideration will be needed on developing alternative versions of the index and how (if at all) to compare between them.

Table 6.3.A   Potential future improvement of indicators measured in the 2019 edition of InCiSE

InCiSE indicator	RAG rating	Potential routes for future development
Attributes		
Integrity	◉	Addition to, or replacement of, existing metrics with non-perception based measures.
Openness	●	n/a
Capabilities	○	Identification of data sources with more recent data and/or more regular update frequency.
Inclusiveness	○	Additional metrics providing objective measurement of ethnic/religious diversity, and metrics providing objective/subjective measurement of inclusion for other under-represented groups (e.g. disability, age, socio-economic background, LGBT).
Functions		
Policymaking	◉	Addition of non-perception based measures, including on themes such as timeliness, accuracy, and use of evidence.
Fiscal & financial management	●	n/a
Regulation	●	n/a
Crisis & risk management	◉	Replacement of the data sourced from the UN's Hyogo Framework for Action monitoring reports as monitoring data from the Sendai Framework becomes available.
HR management	◉	Identification of data to measure additional themes such as skills gaps/talent deployment, quality of learning and development, and level of satisfaction with HR services.
Procurement	◉	Additional themes such as value for money and the capabilities of procurement officials.
Tax administration	◉	Additional themes such as preventing tax evasion.
Digital services	◉	Identification of non-perception based measures, including average transaction time, up-time of systems, proportion of government services available online.

● High data quality   ◉ Medium data quality   ○ Low data quality

References

This chapter lists the data sources used by the InCiSE Index, the software packages used in the analysis and computation of the InCiSE Index, and published works and other materials referred to in this report.

Data sources

Bertelsmann Stiftung (2018) **Sustainable Governance Indicators 2018**, Gütersloh, Germany: Bertelsmann Stiftung, http://www.sgi-network.org

Dahlström C, Teorell J, Dahlberg S, Hartmann F, Lindberg A and Nistotskaya M (2015) **The QoG Expert Survey Dataset II**, Gothenburg: University of Gothenburg, The Quality of Government Institute, https://qog.pol.gu.se/data/datadownloads/qogexpertsurveydata

DIGIWHIST (2018) **Opentender**, Hungary: Government Transparency Institute, https://opentender.eu

European Commission (2018) **eGovernment Benchmark Report 2018**, Brussels: European Commission, https://ec.europa.eu/digital-single-market/en/news/egovernment-benchmark-2018-digital-efforts-european-countries-are-visibly-paying

International Budget Partnership (2018) **Open Budget Survey 2017**, Washington DC: International Budget Partnership, http://survey.internationalbudget.org

ILO (2018) **ILO Statistics**, Geneva: International Labour Organization, https://www.ilo.org/ilostat/

OECD (2013) **Government at a Glance 2013**, Paris: Organisation for Economic Cooperation and Development, https://doi.org/10.1787/gov_glance-2013-en

OECD (2015) **Government at a Glance 2015**, Paris: Organisation for Economic Cooperation and Development, https://doi.org/10.1787/gov_glance-2015-en

OECD (2016) **Dataset on the Governance of Critical Risks**, Paris: Organisation for Economic Cooperation and Development, https://qdd.oecd.org/subject.aspx?Subject=GOV_RISK

OECD (2016) **Dataset on Public Procurement**, Paris: Organisation for Economic Cooperation and Development, https://qdd.oecd.org/subject.aspx?Subject=GOV_PUBPRO_2016

OECD (2017) **Government at a Glance 2017**, Paris: Organisation for Economic Cooperation and Development, https://doi.org/10.1787/gov_glance-2017-en

OECD (2017) **Tax Administration Comparative Information Series**, Paris: Organisation for Economic Cooperation and Development, https://qdd.oecd.org/subject.aspx?Subject=TAS

OECD (2018) **Composite Indicators of Regulatory Policy and Governance**, Paris: Organisation for Economic Cooperation and Development, https://stats.oecd.org/Index.aspx?QueryId=85336

OECD (2018) **Survey of Adult Skills (Programme for International Assessment of Adult Competencies)**, Paris: Organisation for Economic Cooperation and Development, http://www.oecd.org/skills/piaac/

Open Knowledge International (2017) **Global Open Data Index**, London: Open Knowledge Foundation, https://index.okfn.org/

Transparency International (2017) **Global Corruption Barometer**, Berlin: Transparency International, http://gcb.transparency.org/

UNDESA (2018) **United Nations E-Government Survey 2018**, New York: United Nations, https://publicadministration.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2018

UNESCO Institute for Statistics (2018) **Population by minimum completed level of education (cumulative)**, Montreal: UNESCO Institute for Statistics, http://data.uis.unesco.org/index.aspx?queryid=168

UNIDRR (2016) **Hyogo Framework for Action National Progress Query Tool**, New York: United Nations, https://www.preventionweb.net/applications/hfa/qbnhfa/

World Bank (2017) **FMIS And Open Budget Data Global Dataset**, Washington DC: The World Bank, https://datacatalog.worldbank.org/dataset/fmis-and-open-budget-data-global-dataset

World Bank (2018) **Doing Business 2019**, Washington DC: The World Bank, http://www.doingbusiness.org/

World Bank (2018) **Worldwide Development Indicators**, Washington DC: The World Bank, https://data.worldbank.org/indicator/

World Economic Forum (2017) **Global Competitiveness Index 2017-18 (Executive Opinion Survey)**, Geneva: World Economic Forum, http://reports.weforum.org/global-competitiveness-index-2017-2018/downloads/

World Justice Project (2018) **Rule of Law Index 2017-18**, Washington DC: World Justice Project, https://worldjusticeproject.org/our-work/wjp-rule-law-index/wjp-rule-law-index-2017%E2%80%932018

World Wide Web Foundation (2017) **Open Data Barometer: 4th Edition**, Washington DC: World Wide Web Foundation, https://opendatabarometer.org/4thedition/

Software packages

Arel-Bundock V, Enevoldsen N and Yetman CJ (2018) "**countrycode**: An R package to convert country names and country codes", Journal of Open Source Software 3(28): 848, https://doi.org/10.21105/joss.00848

Bescond D (2018) **Rilostat**: ILO Open Data via Ilostat Bulk Download Facility or SDMX Web Service, R package version 0.2.1, https://CRAN.R-project.org/package=Rilostat

Caro DH and Biecek P (2017) "**intsvy**: An R Package for Analyzing International Large-Scale Assessment Data" Journal of Statistical Software 81(7): 1-44, http://doi.org/10.18637/jss.v081.i07

Firke S (2018) **janitor**: Simple Tools for Examining and Cleaning Dirty Data, R package version 1.1.1, https://CRAN.R-project.org/package=janitor

Henry L and Wickham H (2017) **purrr**: Functional Programming Tools, R package version 0.2.4, https://CRAN.R-project.org/package=purrr

Milton Bache S and Wickham H (2014) **magrittr**: A Forward-Pipe Operator for R, R package version 1.5, https://CRAN.R-project.org/package=magrittr

Müller K and Wickham H (2017) **tibble**: Simple Data Frames, R package version 1.3.4, https://CRAN.R-project.org/package=tibble

R Core Team (2016) **R**: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/

van Buuren S and Groothuis-Oudshoorn K (2011) '**mice**: Multivariate Imputation by Chained Equations in R', Journal of Statistical Software 45(3):1-67, http://www.jstatsoft.org/v45/i03/

Wickham H (2016) **ggplot2**: Elegant Graphics for Data Analysis, New York: Springer-Verlag, http://ggplot2.org

Wickham H (2016) **rvest**: Easily Harvest (Scrape) Web Pages, R package version 0.3.2, https://CRAN.R-project.org/package=rvest

Wickham H (2017) **scales**: Scale Functions for Visualization, R package version 0.5.0, https://CRAN.R-project.org/package=scales

Wickham H (2017) **stringr**: Simple, Consistent Wrappers for Common String Operations, R package version 1.2.0, https://CRAN.R-project.org/package=stringr

Wickham H and Bryan J (2017) **readxl**: Read Excel Files, R package version 1.0.0, https://CRAN.R-project.org/package=readxl

Wickham H, Francois R, Henry L and Müller K (2017) **dplyr**: A Grammar of Data Manipulation, R package version 0.7.4, https://CRAN.R-project.org/package=dplyr

Wickham H and Henry L (2017) **tidyr**: Easily Tidy Data with 'spread()' and 'gather()' Functions, R package version 0.7.2, https://CRAN.R-project.org/package=tidyr

Wickham H, Hester J and Francois R (2017) **readr**: Read Rectangular Text Data, R package version 1.1.1, https://CRAN.R-project.org/package=readr

Wickham H and Miller E (2017) **haven**: Import and Export 'SPSS', 'Stata' and 'SAS' Files, R package version 1.1.0, https://CRAN.R-project.org/package=haven

Published works and other materials

Baubion C (2013) *OECD Risk Management: Strategic Crisis Management*, OECD Working Papers on Public Governance Number 23, Paris: OECD, https://doi.org/10.1787/5k41rbd1lzr7-en

Boviard T and Löffler E (2003) *Public Management and Governance*, London: Routledge

Christensen T, Fimreite AL and Lægried P (2011) 'Crisis Management: The Perceptions of Citizens and Civil Servants in Norway', *Administration and Society* 43(5): 561–594

Fukyyama F (2013) 'What is Governance?' *Governance: An International Journal of Policy, Administration and Institutions* 26(3): 347–288

Goujon A, KC S, Speringer M, Barakat B, Potancoková M, Eder J, Striessnig E, Bauer R and Lutz W (2016) 'A Harmonized Dataset On Global Educational Attainment Between 1970 And 2060 – An Analytical Window Into Recent Trends And Future Prospects In Human Capital Development', *Journal of Demographic Economics* 82(3): 315–363.

Gregory M and Upson M (2018) *RAP Companion: Automating the production of statistical reports using DataOps principles [Online]*, UK Government Digital Service, https://ukgovdatascience.github.io/rap_companion/ [last accessed 2 March 2018]

Holt J and Manning N (2014) 'Fukyama is right about measuring state quality: now what?', *Governance: An International Journal of Policy, Administration and Institutions* 27(4): 717–728.

International Civil Service Commission (2002) *Standards of conduct for the international civil service*, New York: United Nations, https://icsc.un.org/resources/pdfs/general/standards.pdf

International Monetary Fund (2016) *Consultation on the IMF Guidelines on Governance Issues [Online]*, https://www.imf.org/external/np/exr/consult/2016/governance/ [last accessed 9 March 2017]

Kaufman D, Kraay A and Zoido-Lobatón P (1999) *Governance Matters*, Policy Research Working Paper 2196, Washington DC: World Bank

Lonti Z and Woods M (2008) *Towards Government at a Glance: Identification of core data and issues related to public sector efficiency*, Paris: OECD, https://doi.org/10.1787/245570167540

Lutz W, Goujon A, KC S and Sanderson WC (2007) *Reconstruction of Populations by Age, Sex and Level of Educational Attainment for 120 Countries for 1970-2000*, IIASA Interim Report, Laxenburg, Austria: International Institute for Applied Systems Analysis, http://pure.iiasa.ac.at/8453

Muriithi K, Jimenez M, Jannin N, Sajid N, Singh S and Sharma S (2015) *Quantifying Governance: an indicator-based approach*, London: LSE/DFID Capstone Team, https://www.gov.uk/dfid-research-outputs/quantifying-governance-an-indicator-based-approach

OECD (2005) *Performance-related Pay Policies for Government Employees*, Paris: OECD https://doi.org/10.1787/9789264007550-en

OECD (2008) *Handbook on Constructing Composite Indicators: Methodology and User Guide, Paris*: OECD, http://oe.cd/composite-indicators-handbook

OECD (2012) *Recommendation of the Council on Regulatory Policy and Governance* (explanatory note), Paris: OECD, http://www.oecd.org/gov/regulatory-policy/49990817.pdf

OECD (2015a) *Government at a Glance 2015*, Paris: OECD, http://dx.doi.org/10/1787/gov_glance_2015-en

OECD (2015b) *Recommendation of the Council on Budgetary Governance* (explanatory note), Paris: OECD, http://www.oecd.org/gov/budgeting/Recommendation-of-the-Council-on-Budgetary-Governance.pdf

Office for Public Management Ltd (OPM) and the Chartered Institute of Public Finance and Accountability (CIPFA) (2004) *The Good Governance Standard for Public Services*, London: Hackney Press Ltd for the Independent Commission on Good Governance in Public Services

UNSD (2018) *Statistical Classifications [Online]*, United Nations, https://unstats.un.org/unsd/class/default.asp [last accessed 2 March 2018]

Wickham H (2015) *R Packages*, Sebastapol CA: O'Reilly Media, http://r-pkgs.had.co.nz

Wickham H and Grolemund G (2017) R for *Data Science*, Sebastapol CA: O'Reilly Media, https://r4ds.had.co.nz/

World Bank (2012) *Indicators of the Strength of Public Management Systems: A key part of the Public Sector results story*, Draft PRMPS Discussion Paper, Washington DC: World Bank, http://siteresources.worldbank.org/PUBLICSECTORANDGOVERNANCE/Resources/285741-1354024300711/ISPMS.pdf

World Bank (2017) *World Development Report 2017*, Washington DC: World Bank

World Trade Organisation (2015) *Government Procurement Agreement: opening markets and promoting good governance [Online]*, http://wto.org/english/thewto_e/20y_e/gpa_brochure2015_e.pdf [last accessed 1 March 2017]

United Nations (1999) *Transparency in Government*. Presentation for ILEA Seminar [Online], http://unpan1.un.org/intradoc/groups/public/documents/un/unpan012062.pdf [last accessed 20 February 2017]

Annex A: Composite metrics

Most metrics in InCiSE are variables taken directly from the source datasets. Some of these metrics are themselves an index or composite score, where this composite is calculated by the source provider the metric is taken "as is" by InCiSE. However, some of the source variables identified for inclusion in InCiSE are binary information (e.g. yes/no questions, or representing categorical data). In the case of inclusiveness the InCiSE model calculates the absolute difference between the composition of central government employees and the composition of the general labour market. This Annex provides details on how the composite metrics calculated by the InCiSE methodology have been produced.

Including binary variables directly into the InCiSE model presents challenges in respect of the usability of the model's results. Firstly, it has the potential for some indicators to become difficult to summarise due to the large number of metrics they would record – in the most extreme case the crisis and risk management indicator would be based on 68 metrics. Secondly, by their nature, binary variables have only two positions, meaning that a country would either score 0 or 1 with no variation between these two positions, limiting the ability to distinguish between relative country performance. To resolve these limitations, the InCiSE model combines binary variables into composite aggregate metrics. These composite metrics have been designed within the InCiSE project's methodology, and have not been developed by the authors/ publishers of the source data. In general, InCiSE maintains conceptual consistency (e.g. the composite metric aggregates information about similar) and source consistency (i.e. a composite metric is an aggregation of variables from the same dataset from the same author/publisher).

Table A.1   Composite metrics in the integrity indicator

InCiSE metric	Source variables	Coding	
Post-employment cooling-off	[OECD] Post-public employment cooling-off: senior civil servants	Cooling-off period for both = 3; Cooling-off for SCS only = 2; Cooling-off for non-SCS only = 1; No cooling-off for both = 0.	
	[OECD] Post-public employment cooling-off: civil servants		
Lobbyist protections	[OECD] Is there an obligation to have a balanced composition of advisory/expert groups?	Yes = 1; No = 0	Sum of variables [Range: 0 to 3]
	[OECD] Are lobbyists allowed to sit in advisory/expert groups in a personal capacity?	Yes = 0; No = 1	
	[OECD] Are corporate executives allowed to sit in advisory/ expert groups in a personal capacity?	Yes = 0; No = 1	
Coverage of whistleblower protections	[OECD] Scope includes: public sector employees	Yes = 1; No = 0	Sum of variables [Range: 0 to 6]
	[OECD] Scope includes: consultants working for the public sector	Yes = 1; No = 0	
	[OECD] Scope includes: suppliers to the public sector	Yes = 1; No = 0	
	[OECD] Scope includes: temporary employees in the public sector	Yes = 1; No = 0	
	[OECD] Scope includes: former public sector employees	Yes = 1; No = 0	
	[OECD] Scope includes: those volunteering for the public sector	Yes = 1; No = 0	

Table A.2   Composite metrics in the inclusiveness indicator

InCiSE metric	Source variables	Coding
Women in central government	[OECD] Women as a proportion of total central government employment	Absolute difference between OECD and ILO variables
	[ILO] Women as a proportion of the labour market	
Women in the public sector	[QoG] Women as a proportion of public sector employment	Absolute difference between QoG and ILO variables
	[ILO] Women as a proportion of the labour market	
Women in top management	[OECD] Women as a proportion of central government senior management positions	Absolute difference between OECD and ILO variables
	[ILO] Women as a proportion of the labour market	
Women in senior government	[QoG] Women as a proportion of senior positions in central government	Absolute difference between QoG and ILO variables
	[ILO] Women as a proportion of the labour market	

Table A.3   Composite metrics in the fiscal and financial indicator

InCiSE metric	Source variables	Coding	
Published public finance data	[WB] Consolidated budget execution results for the public sector?	Yes =1; No = 0	Sum of variables [Range: 0 to 10]
	[WB] Sector analysis?	Yes =1; No = 0	
	[WB] Regional analysis?	Yes =1; No = 0	
	[WB] Gender analysis?	Yes =1; No = 0	
	[WB] Budget analysis with special emphasis towards children and youth?	Yes =1; No = 0	
	[WB] Debt data?	Yes =1; No = 0	
	[WB] Foreign aid data?	Yes =1; No = 0	
	[WB] Fiscal data on sub-national/ local governments and municipalities?	Yes =1; No = 0	
	[WB] Financial statements?	Yes =1; No = 0	
	[WB] Public procurement and contracts for the whole government?	Yes =1; No = 0	

Table A.4   Composite metrics in the tax administration indicator

InCiSE metric	Source variables	Coding
Collection cost	[OECD] Total recurrent budget	Budget as a proportion of net revenue
	[OECD] Net revenue	
Tax debt	[OECD] Total tax debt at year end	Tax debt as a proportion of net revenue
	[OECD] Net revenue	
Online filing: personal tax	[OECD] Personal income tax returns filed online	Online returns as a proportion of total returns
	[OECD] Total personal income tax returns	
Online filing: corporate tax	[OECD] Corporation tax returns filed online	Online returns as a proportion of total returns
	[OECD] Total corporation tax returns	

Table A.5   Composite metrics in the procurement indicator

InCiSE metric	Source variables	Coding	
E-procurement functions	[OECD] Publishing procurement plans	In a national e-procurement system = 1; Only in some specific entities = 0.5; No = 0. If marked as both national and specific systems, then code as 1.	Sum of variables [Range: 0 to 9]
	[OECD] Publication of opportunities		
	[OECD] Announcing tenders		
	[OECD] Online catalogue		
	[OECD] Provision of tender documents		
	[OECD] E-submission of bids		
	[OECD] E-reverse auctions		
	[OECD] Notification of award		
	[OECD] E-submission of invoices		
Role of central purchasing body	[OECD] CPBs award framework agreements or other consolidated instruments, from which CAs then order	Yes = 1; No = 0	Sum of variables [Range: 0 to 4]
	[OECD] CPBs act as CAs aggregating demand and purchasing	Yes = 1; No = 0	
	[OECD] CPBs co-ordinate training for public officials in charge of public procurement	Yes = 1; No = 0	
	[OECD] CPBs establish policies for CAs	Yes = 1; No = 0	
Access for SMEs	[OECD] Specific legislative provision or policy (e.g. set-aside, bid preferences) is in place to encourage the participation of SMEs in procurement	Yes = 1; No = 0	Sum of categories [Range: 0 to 6]
	[OECD] A specific unit specialized in SMEs is in place at the central government level	Yes = 1; No = 0	
	[OECD] Training and workshops are carried out for SMEs	Yes = 1; No = 0	
	[OECD] Documentation or guidance focused on SMEs is available online.	Yes = 1; No = 0	
	[OECD] Division into lots of the contract	Yes = 1; No = 0	
	[OECD] Administrative procedures are simplified for SMEs to participate in tenders (A_CB_1092551)	Yes = 1; No = 0	

Table A.6   Composite metrics in the crisis and risk indicator

InCiSE metric	Source variables	Coding	
Approach	[OECD] Does your government have a national strategy for the management of critical risks?	Yes = 1; No = 0	Sum of variables [Range: 0 to 7]
	[OECD] Does your government's national strategy adopt an all-hazards approach to risk?	Yes = 1; No = 0	
	[OECD] Does your government have an institution that is assigned leadership at the national level for the management of critical risks	Yes = 1; No = 0	
	[OECD] Does the lead institution on the management of critical risks report to the centre of government?	Yes = 1; No = 0	
	[OECD] Does the lead institution prepare a report on its functions to the Head of Government and/or a Cabinet level minister?	Yes = 1; No = 0	
	[OECD] Does the institution consult with a variety of stakeholders in the policy-formulation process for the management of critical risks?	Yes = 1; No = 0	
	[OECD] Does your government have a mechanism for monitoring unexpected events in order to quickly build situation awareness about critical risks once they actually occur?	Yes = 1; No = 0	
Lead institution functions	[OECD] Lead institution functions: design/ formulate risk management policies	Yes = 1; No = 0	Sum of categories [Range: 0 to 11]
	[OECD] Lead institution functions: set priorities and allocate resources accordingly	Yes = 1; No = 0	
	[OECD] Lead institution functions: set performance targets	Yes = 1; No = 0	
	[OECD] Lead institution functions: provide incentives for policy implementation	Yes = 1; No = 0	
	[OECD] Lead institution functions: monitor policy implementation	Yes = 1; No = 0	
	[OECD] Lead institution functions: evaluate policy implementation	Yes = 1; No = 0	
	[OECD] Lead institution functions: disseminate results of evaluation to the public	Yes = 1; No = 0	
	[OECD] Lead institution functions: promote policy coherence across government departments	Yes = 1; No = 0	
	[OECD] Lead institution functions: address competing policy objectives	Yes = 1; No = 0	
	[OECD] Lead institution functions: coordinate actions across central and local levels of government	Yes = 1; No = 0	
	[OECD] Lead institution functions: coordinate cooperation between government and non-governmental entities	Yes = 1; No = 0	
Multi-hazard assessment	[UN] PA2-C1: Multi-hazard risk assessment	Yes = 1; No = 0	Sum of variables [Range: 0 to 5]
	[UN] PA2-C1: Gender disaggregated vulnerability and capacity assessments	Yes = 1; No = 0	
	[UN] PA2-C1: Agreed national standards for multi hazard risk assessments	Yes = 1; No = 0	
	[UN] PA2-C1: Common format for risk assessment	Yes = 1; No = 0	
	[UN] PA2-C1: Is future/probable risk assessed?	Yes = 1; No = 0	

Table A.6 (continued)

InCiSE metric	Source variables	Coding	
Risk monitoring	[UN] PA2-C2: Are disaster losses and hazards systematically reported, monitored and analyzed?	Yes = 1; No = 0	Sum of variables [Range: 0 to 4]
	[UN] PA2-C2: Disaster loss databases exist and are regularly updated	Yes = 1; No = 0	
	[UN] PA2-C2: Reports generated and used in planning by finance, planning and sectoral line ministries (from the disaster databases/ information systems)	Yes = 1; No = 0	
	[UN] PA2-C2: Hazards are consistently monitored across localities and territorial boundaries	Yes = 1; No = 0	
Risk management capability	[OECD] Does your government undertake efforts to develop risk anticipation capacity	Yes = 1; No = 0	Sum of variables [Range: 0 to 5]
	[OECD] Does your government's national strategy for the management of critical risks promote measures to enhance risk prevention and mitigation?	Yes = 1; No = 0	
	[OECD] Does your government have a critical infrastructure protection programme (CIP)?	Yes = 1; No = 0	
	[OECD] Are inter-agency cooperation mechanisms built into your government's crisis management system?	Yes = 1; No = 0	
	[OECD] Does your government encourage the private sector to take steps to ensure business continuity?	Yes = 1; No = 0	
Preparedness	[UN] PA5-C1: Are future disaster risks anticipated through scenario development and aligned preparedness planning?	Yes = 1; No = 0	Sum of variables [Range: 0 to 5]
	[UN] PA5-C1: Are there national programmes or policies for disaster preparedness, contingency planning and response?	Yes = 1; No = 0	
	[UN] PA5-C1: The institutional mechanisms exist for the rapid mobilisation of resources in a disaster, utilising civil society and the private sector; in addition to public sector support.	Yes = 1; No = 0	
	[UN] PA5-C1: Preparedness plans are regularly updated based on future risk scenarios	Yes = 1; No = 0	
	[UN] PA5-C2: Risk management/contingency plans for continued basic service delivery	Yes = 1; No = 0	
Disaster spending appraisal	[UN] PA4-C3: Are the costs and benefits of DRR incorporated into the planning of public investment?	Yes = 1; No = 0	Sum of variables [Range: 0 to 6]
	[UN] PA4-C6: Are the impacts of disaster risk that are created by major development projects assessed?	Yes = 1; No = 0	
	[UN] PA4-C6: Are cost/benefits of disaster risk taken into account in the design and operation of major development projects?	Yes = 1; No = 0	
	[UN] PA4-C6: Impacts of disaster risk taken account in Environment Impact Assessment (EIA)	Yes = 1; No = 0	
	[UN] PA4-C6: By national and sub-national authorities and institutions	Yes = 1; No = 0	
	[UN] PA4-C6: By international development actors	Yes = 1; No = 0	

InCiSE metric	Source variables	Coding	
International cooperation	[UN] PA2-C4: Does your country participate in regional or sub-regional actions to reduce disaster risk?	Yes = 1; No = 0	Sum of variables [Range: 0 to 6]
	[UN] PA2-C4: Establishing and maintaining regional hazard monitoring	Yes = 1; No = 0	
	[UN] PA2-C4: Regional or sub-regional risk assessment	Yes = 1; No = 0	
	[UN] PA2-C4: Regional or sub-regional early warning	Yes = 1; No = 0	
	[UN] PA2-C4: Establishing and implementing protocols for transboundary information sharing	Yes = 1; No = 0	
	[UN] PA2-C4: Establishing and resourcing regional and sub-regional strategies and frameworks	Yes = 1; No = 0	
Risk communications	[OECD] Does your government encourage a whole-of-society approach to risk communication?	Yes = 1; No = 0	Sum of variables [Range: 0 to 3]
	[OECD] Has your government communicated the results of any such evaluations to the public in the past?	Yes = 1; No = 0	
	[OECD] Does your government make information that is used for the assessment of critical risks available to the public?	Yes = 1; No = 0	
Early warning systems	[UN] PA2-C3: Do risk prone communities receive timely and understandable warnings of impending hazard events?	Yes = 1; No = 0	Sum of variables [Range: 0 to 3]
	[UN] PA2-C3: Communication systems and protocols used and applied	Yes = 1; No = 0	
	[UN] PA2-C3: Active involvement of media in early warning dissemination	Yes = 1; No = 0	
Risk evaluation and research	[OECD] Has your government conducted a post-disaster evaluation of policies that are designed to support the management of critical risks within the last three years?	Yes = 1; No = 0	Sum of variables [Range: 0 to 3]
	[OECD] Have the results from such evaluations been used in the design of revised risk management policies?	Yes = 1; No = 0	
	[OECD] Does your government provide support for scientific research that is meant to improve policies for the management of critical risks?	Yes = 1; No = 0	
Post-disaster assessment	[UN] PA5-C4: Damage and loss assessment methodologies and capacities available	Yes = 1; No = 0	Sum of variables [Range: 0 to 3]
	[UN] PA5-C4: Post-disaster needs assessment methodologies	Yes = 1; No = 0	
	[UN] PA5-C4: Post-disaster needs assessment methodologies include guidance on gender aspects	Yes = 1; No = 0	

Annex B: Sensitivity analysis – detailed results

This Annex provides detailed results from the sensitivity analysis described in Chapter 5. Each table includes the index score and rank for each of the 38 countries included in the 2019 InCiSE Index results for each of the sensitivity tests carried out alongside the results of the 2019 index results.

Table B.1 shows the results of the sensitivity tests varying country coverage:

- Using a data quality assessment threshold of 0.55 to determine country inclusion;
- Using a data quality assessment threshold of 0.6 to determine country inclusion;
- Using a threshold of 75% of the available data to determine country inclusion;
- Using only the countries included in the 2017 Pilot edition of the index.

Table B.2 shows the results of the sensitivity tests varying the reference date:

- Excluding the capabilities indicator;
- Excluding the capabilities indicator and adjusting the weighting accordingly;
- Using only data with a reference year of 2015 or later;
- Using only data with a reference year of 2016 or later.

Table B.3 shows the results of the sensitivity tests using alternative weighting:

- Using a 50:50 split for the equal-share and data-quality based weighting;
- Using only equal indicator weights (i.e. all indicator weights equal 1/12);
- Using only indicators weights based on the data quality assessment results;
- Not applying any within-indicator weights;
- Calculating the index as a sum of all metrics.

Table B.4 shows the results of the sensitivity tests adjusting the base data:

- Ranking the metrics before imputation;
- Rescaling the metrics before imputation;
- Standardising the metrics before imputation.

Table B.5 shows the results of the sensitivity tests using alternative imputation methods:

- Using an 'all-in-one' approach for imputation of missing data;
- Using the 'midas touch' method for imputation of missing data;
- Using the 'random forests' method for imputation of missing data;
- Replacing missing data with the mean of observed values.

Table B.1   Sensitivity tests varying country coverage

Country	Index score					Country rank				
	2019 results	DQA ≥ 0.55	DQA ≥ 0.6	75 % of data	2017 group	2019 results	DQA ≥ 0.55	DQA ≥ 0.6	75 % of data	2017 group
GBR	1.000	1.000	1.000	1.000	1.000	1	1	1	1	1
NZL	0.980	0.987	0.987	0.992	0.985	2	2	2	2	2
CAN	0.916	0.907	0.902	0.906	0.898	3	3	3	3	3
FIN	0.883	0.887	0.881	0.886	0.879	4	4	4	4	4
AUS	0.863	0.859	0.858		0.859	5	5	5		5
DNK	0.832	0.854	0.847	0.850	0.843	6	6	6	5	6
NOR	0.830	0.832	0.830	0.831	0.828	7	7	7	6	7
NLD	0.794	0.792	0.790	0.790	0.786	8	8	8	7	8
KOR	0.785	0.781	0.779	0.777	0.773	9	9	9	8	10
SWE	0.785	0.773	0.775	0.775	0.775	10	10	10	9	9
USA	0.765	0.759			0.758	11	11			11
EST	0.674	0.675	0.674	0.672	0.671	12	12	11	10	12
CHE	0.650	0.639	0.635	0.641	0.635	13	13	12	11	13
IRL	0.625	0.623	0.617	0.619	0.611	14	16	15	14	16
FRA	0.619	0.626	0.627	0.627	0.628	15	15	13	12	14
AUT	0.617	0.626	0.621	0.624	0.620	16	14	14	13	15
ESP	0.599	0.596	0.589	0.586	0.590	17	17	16	15	17
MEX	0.507	0.500	0.504		0.505	18	19	17		18
DEU	0.505	0.502	0.503	0.504	0.502	19	18	18	16	19
LTU	0.487	0.490	0.483	0.484		20	20	19	18	
BEL	0.485	0.483	0.475	0.482	0.473	21	22	21	19	21
JPN	0.472	0.485	0.481	0.488	0.479	22	21	20	17	20
LVA	0.466	0.471	0.463	0.462		23	23	22	20	
CHL	0.454	0.451	0.450	0.452	0.446	24	24	23	21	22
ITA	0.419	0.431	0.428	0.424	0.427	25	25	24	22	23
SVN	0.369	0.359	0.358	0.360	0.359	26	26	25	23	24
ISR	0.315	0.322				27	27			
POL	0.282	0.277	0.274	0.281	0.274	28	28	26	24	25
PRT	0.259	0.260	0.260	0.259	0.258	29	29	27	25	27
CZE	0.245	0.260	0.260	0.257	0.259	30	30	28	26	26
ISL	0.228	0.233				31	31			
TUR	0.189	0.181	0.184	0.176	0.182	32	32	29	27	28
SVK	0.172	0.178	0.175	0.171	0.175	33	33	30	28	29
BGR	0.147					34				
HRV	0.140					35				
ROU	0.127					36				
GRC	0.107	0.103	0.098	0.092	0.098	37	34	31	29	30
HUN	0.000	0.000	0.000	0.000	0.000	38	35	32	30	31
Mean Absolute Error (MAE)		0.006	0.006	0.007	0.007					

Table B.2   Sensitivity tests varying reference year

Country	Index score					Country rank				
	2019 results	Excl. CAP	Excl. CAP & rewght	2015-18 data	2016-18 data	2019 results	Excl. CAP	Excl. CAP & rewght	2015-18 data	2016-18 data
GBR	1.000	1.000	1.000	1.000	1.000	1	1	1	1	1
NZL	0.980	0.955	0.955	0.949	0.912	2	2	2	2	2
CAN	0.916	0.906	0.906	0.895	0.834	3	3	3	3	3
FIN	0.883	0.873	0.873	0.881	0.830	4	5	5	4	4
AUS	0.863	0.875	0.875	0.869	0.810	5	4	4	5	5
DNK	0.832	0.817	0.817	0.789	0.775	6	7	7	9	8
NOR	0.830	0.836	0.836	0.843	0.791	7	6	6	6	6
NLD	0.794	0.802	0.802	0.790	0.758	8	8	8	8	9
KOR	0.785	0.799	0.799	0.784	0.711	9	9	9	10	11
SWE	0.785	0.791	0.791	0.802	0.787	10	10	10	7	7
USA	0.765	0.740	0.741	0.735	0.731	11	11	11	11	10
EST	0.674	0.664	0.665	0.639	0.594	12	12	12	12	17
CHE	0.650	0.653	0.653	0.639	0.597	13	14	14	13	14
IRL	0.625	0.622	0.623	0.619	0.594	14	16	16	15	15
FRA	0.619	0.655	0.656	0.628	0.636	15	13	13	14	12
AUT	0.617	0.635	0.636	0.606	0.594	16	15	15	16	16
ESP	0.599	0.608	0.609	0.593	0.602	17	17	17	17	13
MEX	0.507	0.519	0.519	0.519	0.523	18	20	20	19	19
DEU	0.505	0.521	0.521	0.506	0.536	19	19	19	21	18
LTU	0.487	0.534	0.534	0.518	0.484	20	18	18	20	20
BEL	0.485	0.506	0.506	0.526	0.434	21	21	21	18	22
JPN	0.472	0.495	0.495	0.482	0.444	22	22	22	22	21
LVA	0.466	0.474	0.474	0.454	0.349	23	23	23	24	26
CHL	0.454	0.463	0.463	0.473	0.396	24	24	24	23	25
ITA	0.419	0.451	0.452	0.442	0.427	25	25	25	25	23
SVN	0.369	0.363	0.363	0.351	0.271	26	26	26	26	28
ISR	0.315	0.332	0.333	0.326	0.401	27	27	27	27	24
POL	0.282	0.268	0.269	0.237	0.083	28	28	28	30	36
PRT	0.259	0.256	0.256	0.238	0.220	29	30	30	29	29
CZE	0.245	0.240	0.240	0.221	0.309	30	31	31	32	27
ISL	0.228	0.238	0.238	0.233	0.218	31	32	32	31	30
TUR	0.189	0.261	0.262	0.244	0.141	32	29	29	28	32
SVK	0.172	0.156	0.156	0.159	0.150	33	34	34	33	31
BGR	0.147	0.153	0.153	0.135	0.110	34	35	35	34	34
HRV	0.140	0.144	0.144	0.118	0.073	35	36	36	36	37
ROU	0.127	0.128	0.128	0.105	0.104	36	37	37	37	35
GRC	0.107	0.158	0.158	0.129	0.113	37	33	33	35	33
HUN	0.000	0.000	0.000	0.000	0.000	38	38	38	38	38
Mean Absolute Error (MAE)		0.015	0.015	0.017	0.045					

Table B.3 Sensitivity tests with alternative approaches to weighting

Country	Index score						Country rank					
	2019 results	50:50	Equal wgt	DQA wgt	No w/in indctr	Sum of metrics	2019 results	50:50	Equal wgt	DQA wgt	No w/in indctr	Sum of metrics
GBR	1.000	1.000	0.999	1.000	1.000	0.949	1	1	2	1	1	2
NZL	0.980	0.970	1.000	0.941	0.941	1.000	2	2	1	2	2	1
CAN	0.916	0.911	0.927	0.895	0.896	0.928	3	3	3	3	3	3
FIN	0.883	0.876	0.895	0.858	0.824	0.897	4	4	4	4	5	4
AUS	0.863	0.859	0.869	0.850	0.864	0.864	5	5	5	5	4	5
DNK	0.832	0.823	0.851	0.797	0.774	0.855	6	7	6	7	8	6
NOR	0.830	0.825	0.839	0.812	0.805	0.851	7	6	7	6	7	7
NLD	0.794	0.787	0.809	0.765	0.745	0.783	8	8	8	10	10	10
KOR	0.785	0.786	0.784	0.786	0.809	0.787	9	9	11	8	6	9
SWE	0.785	0.782	0.790	0.775	0.763	0.805	10	10	9	9	9	8
USA	0.765	0.754	0.787	0.722	0.724	0.781	11	11	10	11	11	11
EST	0.674	0.667	0.687	0.649	0.657	0.740	12	12	12	12	12	12
CHE	0.650	0.647	0.656	0.637	0.613	0.639	13	13	13	13	14	14
IRL	0.625	0.616	0.641	0.593	0.598	0.601	14	15	14	15	16	17
FRA	0.619	0.618	0.620	0.615	0.635	0.609	15	14	16	14	13	16
AUT	0.617	0.608	0.635	0.583	0.607	0.649	16	16	15	16	15	13
ESP	0.599	0.593	0.611	0.576	0.589	0.611	17	17	17	17	17	15
MEX	0.507	0.511	0.497	0.524	0.497	0.412	18	18	19	18	19	24
DEU	0.505	0.503	0.507	0.499	0.514	0.525	19	19	18	19	18	18
LTU	0.487	0.489	0.483	0.494	0.482	0.470	20	20	21	20	20	22
BEL	0.485	0.482	0.491	0.473	0.477	0.488	21	21	20	21	21	21
JPN	0.472	0.471	0.475	0.467	0.438	0.503	22	22	23	22	24	19
LVA	0.466	0.460	0.479	0.442	0.447	0.498	23	23	22	24	23	20
CHL	0.454	0.451	0.459	0.444	0.461	0.440	24	24	24	23	22	23
ITA	0.419	0.419	0.418	0.420	0.436	0.368	25	25	25	25	25	26
SVN	0.369	0.360	0.388	0.333	0.338	0.394	26	26	26	26	26	25
ISR	0.315	0.309	0.326	0.294	0.314	0.334	27	27	27	27	27	27
POL	0.282	0.276	0.293	0.261	0.236	0.291	28	28	28	28	29	28
PRT	0.259	0.253	0.273	0.235	0.241	0.289	29	29	29	29	28	29
CZE	0.245	0.240	0.253	0.228	0.174	0.254	30	30	30	30	32	31
ISL	0.228	0.223	0.239	0.207	0.218	0.270	31	31	31	32	31	30
TUR	0.189	0.194	0.177	0.209	0.231	0.063	32	32	32	31	30	35
SVK	0.172	0.170	0.175	0.166	0.107	0.143	33	33	33	33	35	32
BGR	0.147	0.146	0.150	0.141	0.129	0.127	34	34	34	34	33	33
HRV	0.140	0.140	0.140	0.140	0.086	0.069	35	35	35	35	36	34
ROU	0.127	0.131	0.120	0.140	0.070	0.050	36	36	36	36	37	36
GRC	0.107	0.111	0.099	0.123	0.123	0.000	37	37	37	37	34	38
HUN	0.000	0.000	0.000	0.000	0.000	0.023	38	38	38	38	38	37
Mean Absolute Error (MAE)		0.004	0.009	0.017	0.026	0.032						

Table B.4 Sensitivity tests adjusting the base data

Country	Index score				Country rank			
	2019 results	Ranked data	Rescale data	Standardise data	2019 results	Ranked data	Rescale data	Standardise data
GBR	1.000	1.000	1.000	1.000	1	1	1	1
NZL	0.980	0.951	0.978	0.979	2	2	2	2
CAN	0.916	0.893	0.927	0.908	3	3	3	3
FIN	0.883	0.883	0.885	0.885	4	4	4	4
AUS	0.863	0.833	0.868	0.864	5	6	5	5
DNK	0.832	0.862	0.824	0.827	6	5	7	7
NOR	0.830	0.789	0.830	0.829	7	10	6	6
NLD	0.794	0.815	0.794	0.783	8	8	8	9
KOR	0.785	0.818	0.767	0.770	9	7	10	11
SWE	0.785	0.802	0.785	0.784	10	9	9	8
USA	0.765	0.775	0.737	0.781	11	11	11	10
EST	0.674	0.680	0.666	0.671	12	13	12	12
CHE	0.650	0.681	0.630	0.625	13	12	15	13
IRL	0.625	0.588	0.631	0.623	14	15	14	14
FRA	0.619	0.632	0.633	0.620	15	14	13	15
AUT	0.617	0.573	0.622	0.612	16	17	16	16
ESP	0.599	0.577	0.598	0.591	17	16	17	17
MEX	0.507	0.461	0.517	0.522	18	23	18	18
DEU	0.505	0.506	0.497	0.488	19	20	19	21
LTU	0.487	0.495	0.490	0.497	20	21	21	20
BEL	0.485	0.514	0.493	0.513	21	19	20	19
JPN	0.472	0.521	0.478	0.484	22	18	23	22
LVA	0.466	0.432	0.461	0.453	23	24	24	24
CHL	0.454	0.465	0.484	0.474	24	22	22	23
ITA	0.419	0.402	0.415	0.412	25	25	25	25
SVN	0.369	0.386	0.366	0.361	26	26	26	26
ISR	0.315	0.322	0.324	0.327	27	27	27	27
POL	0.282	0.242	0.276	0.273	28	29	28	28
PRT	0.259	0.223	0.252	0.238	29	31	31	31
CZE	0.245	0.275	0.257	0.251	30	28	29	30
ISL	0.228	0.230	0.254	0.263	31	30	30	29
TUR	0.189	0.185	0.179	0.190	32	32	32	32
SVK	0.172	0.170	0.177	0.162	33	33	33	33
BGR	0.147	0.122	0.135	0.113	34	35	35	35
HRV	0.140	0.131	0.151	0.139	35	34	34	34
ROU	0.127	0.107	0.102	0.090	36	36	36	36
GRC	0.107	0.072	0.066	0.075	37	37	37	37
HUN	0.000	0.000	0.000	0.000	38	38	38	38
Mean Absolute Error (MAE)		0.021	0.010	0.011				

Table B.5   Sensitivity tests using alternative imputation methods

Country	Index score					Country rank				
	2019 results	All-in-one	Midas touch	Random forests	Mean	2019 results	All-in-one	Midas touch	Random forests	Mean
GBR	1.000	0.989	1.000	1.000	1.000	1	2	1	1	1
NZL	0.980	1.000	0.975	0.981	0.966	2	1	2	2	2
CAN	0.916	0.862	0.902	0.891	0.867	3	5	3	3	3
FIN	0.883	0.881	0.884	0.874	0.850	4	3	4	4	4
AUS	0.863	0.872	0.807	0.854	0.821	5	4	7	5	6
DNK	0.832	0.833	0.838	0.837	0.808	6	7	5	6	7
NOR	0.830	0.834	0.831	0.831	0.824	7	6	6	7	5
NLD	0.794	0.791	0.794	0.766	0.762	8	9	8	9	9
KOR	0.785	0.738	0.764	0.755	0.728	9	10	10	10	10
SWE	0.785	0.799	0.792	0.785	0.788	10	8	9	8	8
USA	0.765	0.691	0.747	0.721	0.699	11	11	11	11	11
EST	0.674	0.639	0.627	0.665	0.590	12	12	13	12	15
CHE	0.650	0.621	0.646	0.613	0.595	13	15	12	14	13
IRL	0.625	0.562	0.608	0.601	0.534	14	17	16	16	17
FRA	0.619	0.629	0.619	0.603	0.605	15	13	14	15	12
AUT	0.617	0.621	0.618	0.614	0.593	16	14	15	13	14
ESP	0.599	0.609	0.598	0.598	0.574	17	16	17	17	16
MEX	0.507	0.460	0.498	0.485	0.482	18	20	18	19	19
DEU	0.505	0.488	0.494	0.484	0.490	19	18	19	20	18
LTU	0.487	0.431	0.472	0.474	0.427	20	22	20	21	22
BEL	0.485	0.465	0.471	0.501	0.447	21	19	21	18	20
JPN	0.472	0.421	0.460	0.452	0.429	22	24	22	23	21
LVA	0.466	0.367	0.414	0.423	0.365	23	26	24	24	25
CHL	0.454	0.447	0.458	0.469	0.419	24	21	23	22	23
ITA	0.419	0.426	0.412	0.405	0.397	25	23	25	25	24
SVN	0.369	0.395	0.368	0.363	0.358	26	25	26	26	26
ISR	0.315	0.246	0.294	0.322	0.269	27	29	27	27	27
POL	0.282	0.302	0.277	0.262	0.260	28	27	28	29	28
PRT	0.259	0.294	0.266	0.241	0.252	29	28	29	31	30
CZE	0.245	0.219	0.227	0.241	0.243	30	30	31	30	31
ISL	0.228	0.207	0.235	0.297	0.254	31	31	30	28	29
TUR	0.189	0.184	0.195	0.197	0.180	32	33	32	32	32
SVK	0.172	0.203	0.170	0.153	0.132	33	32	33	33	34
BGR	0.147	0.138	0.144	0.095	0.129	34	35	35	36	35
HRV	0.140	0.144	0.159	0.145	0.169	35	34	34	34	33
ROU	0.127	0.095	0.128	0.101	0.121	36	37	36	35	36
GRC	0.107	0.119	0.081	0.075	0.066	37	36	37	37	37
HUN	0.000	0.000	0.000	0.000	0.000	38	38	38	38	38
Mean Absolute Error (MAE)		0.026	0.011	0.018	0.032					

InCiSE