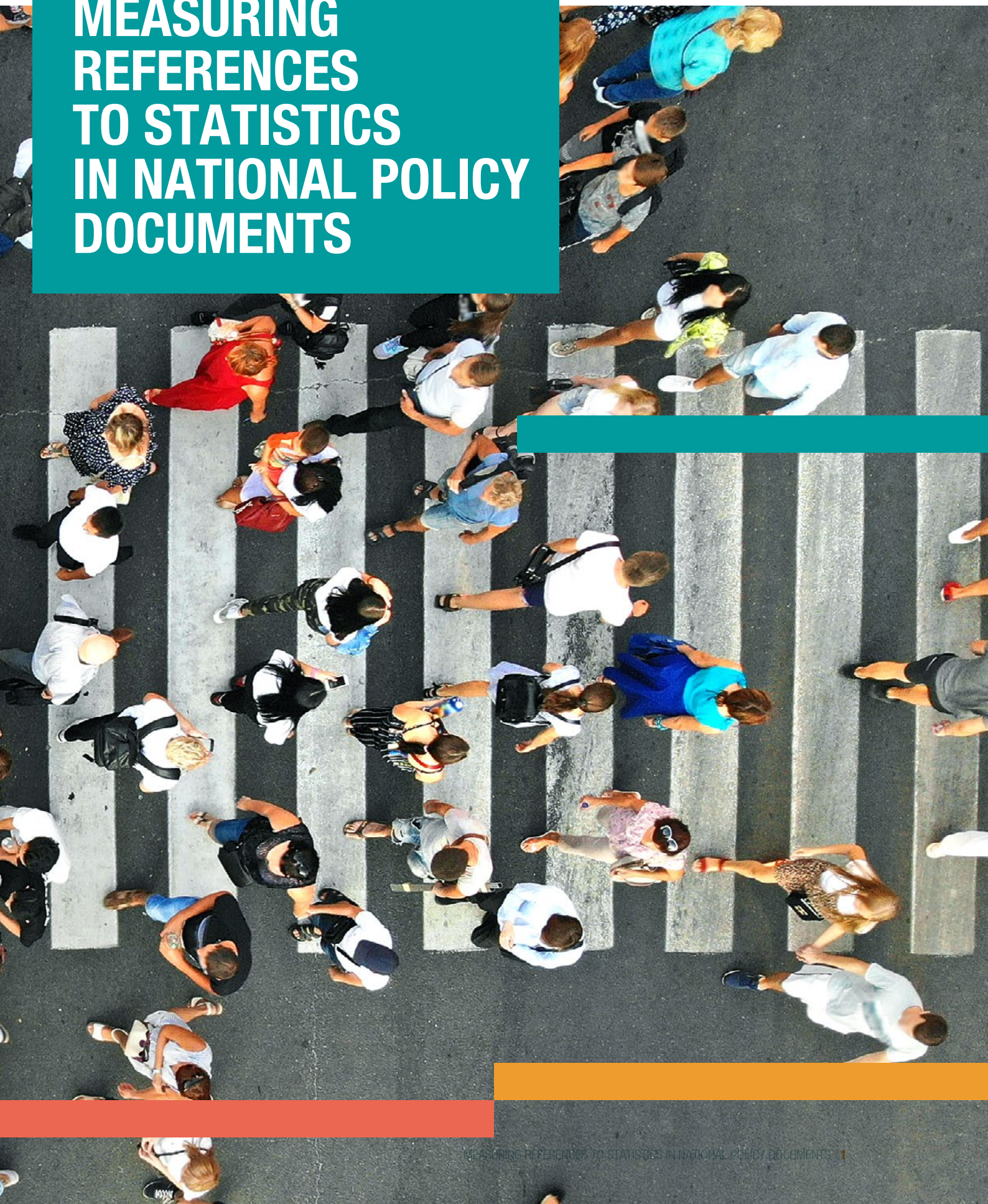
PARiS 21
Partnership in statistics for development in the 21st century

MEASURING REFERENCES TO STATISTICS IN NATIONAL POLICY DOCUMENTS



The report was prepared by the Secretariat of the Partnership in Statistics for Development in the 21st Century (PARIS21).

PARIS21 promotes the better use and production of statistics throughout the developing world. Since its establishment in 1999, PARIS21 has successfully developed a worldwide network of statisticians, policy makers, analysts, and development practitioners committed to evidence-based decision making. With the main objective to achieve national and international development goals and reduce poverty in low and middle income countries, PARIS21 facilitates statistical capacity development, advocates for the integration of reliable data in decision making, and co-ordinates donor support to statistics.

Acknowledgements:

Yu Tian drafted this paper under the supervision of Rajiv Ranjan. We cannot express enough thanks to Rolando Avendano (ADB) and Barbara Baredes (OECD), who contributed significantly to the development of the methodology and early drafts of the paper.

The thematic development on gender statistics could not have been accomplished without the contribution from Lauren Harrison, Liliana Suchodolska and Paz Patiño (PARIS21). We want to express our deep gratitude to UN Women for supporting the thematic development. We are also grateful to all of our expert meeting participants.

We also express sincere gratitude to El Iza Mohamedou (OECD), François Fonteneau and Johannes Jütting (PARIS21) for their continued support and encouragement.

This paper benefited greatly from the following reviewers and their expertise and insights: Aude Bicquelet-Lock (Royal Town Planning Institute), Radoslav Zubek (University of Oxford), Enzo Loner (University of Trento), Jessamyn Encarnacion (UN Women), Lorenz Noe (ODW), Mirta Lourenço (UNESCO) and Stella Nassolo (Uganda Bureau of Statistics), Amie Joof (FAMEDEV), Andrea Arzaba (Global Investigative Journalism Network) and Vanessa Berry-Chatelain (OECD).

We are also grateful for the thoughtful feedback and comments from Rajiv Ranjan, Sasha Ramirez–Hughes, Julia Schmidt, Lauren Harrison, Liliana Suchodolska and Paz Patiño (PARIS21).

The paper is designed by Karamelaki Creatives.

Please cite this publication as: PARIS21 (2021), “Measuring References to Statistics in National Policy Documents”, Paris.

Available at: http://paris21.org/uos

Contact the PARIS21 Secretariat at contact@paris21.org
Visit www.PARIS21.org
Follow PARIS21
@contactparis21
/company/paris21/

ABSTRACT

Evidence-based decision making is essential to policy design and, more broadly, for achieving the Sustainable Development Goals. Policymakers are expected to analyse and use evidence to design, validate and monitor their action plan. This paper develops and reports on a methodology that measures the references to and critical engagement with statistical concepts, in national policy documents (mostly national development plans) and poverty reduction strategies. By analysing over 200 policy documents for quality, the indicator constructed by this methodology showed both robustness and interpretability.

CONTENTS

FIGURES

TABLES

BOXES

1

1. INTRODUCTION

Official statistics and good governance go hand in hand. The ultimate goal of official and publicly available data is enabling governments and policymakers to make informed decisions and hold representative bodies accountable, as acknowledged in global initiatives including the Cape Town Global Action Plan for Sustainable Development Data.[1] The use of evidence is relevant for public policy decisions, ranging from executive orders to regional and national development plans and, more broadly, for achieving the Sustainable Development Goals (SDGs). For this, governments are being called to analyse empirical evidence with the purpose of identifying issues that require action. Subsequently, they must design and monitor potential interventions, monitor the implementation of policies, and evaluate their impact. The evidence is also used to share information on progress with stakeholders in order that an assessment can be made as to whether commitments are being honoured.

The literature has identified two main channels through which statistics have an effect on policymaking. The first is efficiency, whereby the availability of indicators allows for the better allocation of resources. Statistics also allow governments to track the deployment of public resources, such as the distribution of public investment. The second channel is public regardedness, that is, the extent to which a policy is designed to promote public welfare. Statistics enable the effect of public policies to be monitored and help to enhance those that are more directly linked to public well-being (Ardanaz, Scartascini and Tommasi, 2010).

Statistical capacity is affected by the political economy and hence the potential use of statistics for better policymaking. These factors include the existence of government entities that require statistics to function, the international demand for information and support to statistical offices, the role of the executive in promoting statistical development, and the presence of non-governmental actors that demand high quality statistics to hold their governments to account (Dargent et. al., 2018: 69). These factors relate to the channels through which statistics affect policymaking (e.g., public regardedness requires interest on the part of international organisations, the civil society and the private sector in monitoring policies).

There is further room to explore the link between the use of evidence in the policy-making process and policy effectiveness. Understanding the role of statistical evidence prior to the policy-making process could facilitate governments gauging the capacity to pursue and attain positive development outcomes in the medium term.

For the purposes of this work, Use of Statistics (henceforth UoS) is defined as the systematic use of statistical knowledge to inform programme design and policy choice, monitor policy implementation, and evaluate policy impact (Russell and Munoz Ayala, 2015; Scott, 2005). With this objective in mind, this paper proposes an improved indicator for measuring the use of statistical evidence in national policy documents from 1994-2017 (Annex 1). It then provides a methodology for constructing the index and compares the relationship between the index and other development outcomes. Finally, it concludes with some reflections on the best strategies to improve the use of data in policy documents.

DISENTANGLING THE USE OF DATA IN NATIONAL DEVELOPMENT PLANNING

As the master plan for national policy, a national development plan analyses a country's objectives and priorities in relation to all sectors in the national economy in response to well-identified national needs. It also proposes and justifies an overall plan in which the role of individual sectors can be seen

[1] *https://unstats.un.org/sdgs/hlg/cape-town-global-action-plan/*

in context. Measuring the use of statistics in national development plans has thus become the most straightforward entry point for analysing the use of statistics in policymaking.

The design and implementation of national public policies is a complex process involving numerous actors and thorough consultations. National policy documents (e.g., National Development Plans, Poverty Reduction Strategy Papers) are outputs of this process that aim to condense the vision and objectives, and to establish practical steps that will guide governments in their efforts to overcome poverty and development bottlenecks in the medium term. Such documents are designed and written as frameworks that set out strategies to, inter alia, promote social development and improve living conditions.

A National Development Plan (henceforth NDP) is a policy framework that outlines a country's strategy to promote social development and improve living conditions across the national territory. Although there is neither an established scope nor structure for the document, the NDP is the formal policy instrument through which a government commits to certain development objectives and delineates how performance will be evaluated. The Ministry of Planning or another Centre-of-Government (CoG) entity is responsible for the consultation process involved in the design, implementation and evaluation of the plan, as well as coordinating actions for achieving the objectives.

Similarly, Poverty Reduction Strategy Papers (PRSP) were conceived as policy documents for low-income countries to (i) outline a poverty-reduction strategy for the medium term and (ii) envision multilateral plans (World Bank and IMF) with the aim of making use of the debt alleviation funds from the Highly Indebted Countries Initiative. By evaluating policies from a pro-poor perspective, the PRSPs are a necessary condition for HIC countries to access conditional financing.

In general, NDPs and PRSPs are focused on 3 to 5 main domains of a country's national development strategy: a) a macroeconomic and fiscal assessment of the country's position over the next 5 to10 years; b) an assessment of poverty and inequality and social policies; c) an assessment of current sectoral and industrial policies; d) an analysis of the international economic context and the country's trade and financial integration strategy; and e) a monitoring and evaluation provisions for tracking progress towards the accomplishment of objectives. The document may also contain budget allocations and cost estimates. With varying levels of detail, national policy documents provide an overview of the current developments in those areas and introduce policies and programmes to improve outcomes. Since the Millennium Declaration was signed in the early 2000s, governments have been inclined to incorporate international commitments into national agendas and articulate them in NDPs and PRSPs. Incrementally, the 2030 agenda, whose progress was monitored through a framework of 231 unique SDG indicators,[2] has permeated and become part of objectives in the national policy documents, both for the definition of objectives and for monitoring purposes.

The first indicator on the use of statistics in policymaking was developed by PARIS21 (2010) following an evaluation that detected the impact of statistical capacity development on NDPs and PRSPs. The indicator was later used as part of the monitoring process of the Busan Action Plan for Statistics. The endorsement of the Cape Town Global Action Plan in 2017 increased the interest from within the data community in adapting the methodology and reflecting on the role of users. Building on the original indicator, a new methodology will seek to capture how decisions reflected in national planning documents are made and supported through statistical evidence (PARIS21, 2015). The updated methodology builds on PARIS21's approach for measuring the use of statistics in PARIS21 (2015) and suggests several methodological improvements to increase the accuracy and relevance of

[2] *The total number of indicators listed in the global indicator framework of SDG indicators is 247. However, twelve indicators repeat under two or three different targets. See https://unstats.un.org/sdgs/indicators/indicators-list*

the indicator. These include an expansion in geographic, temporal and thematic coverage; a refined methodology for the selection of keywords; and the inclusion of "levels of use".

These methodological improvements were complemented by additional work on specific sectoral applications. In 2019, PARIS21 collaborated with FAO to expand the scope of work from generic documents to food and agriculture policy documents. In addition, as part of its collaboration with UN Women, PARIS21 is exploring the use of gender statistics (including disaggregation by sex and other variables) in national policy documents and sectoral gender policy documents to identify patterns in gender-sensitive approaches to data use and policy design (Annex 1).

Based on the previous work, the authors of this paper redeveloped the methodology to measure references to statistics in various policy documents. A wider range of documents were selected, extensive and robust keyword lists were created, and new weighting methods were applied. This methodology allows users, for drafting, enhanced policymaking or other academic purposes, to analyse their policy documents in an automatic, customisable and efficient way.

LITERATURE REVIEW

The use of text mining and machine learning techniques has become increasingly relevant in different areas of economic analysis. To analyse policymaking, Hansen, McMahon and Prat (2018) explore the use of computational linguistics algorithms for analysing how transparency affects the deliberation of monetary policymakers and their decision-making process. Meanwhile, Dupriez (2018) applies machine learning algorithms for predicting poverty status using households' survey information and Pincet et al. (2019) use machine learning techniques to analyse information on the OECD Credit Reporting System (CRS) and its contribution to the SDGs.

The effect of social norms has been analysed through machine learning. Parthasarathy, Rao and Palaniswamy (2017) used Natural Language Processing (NLP) methods on a corpus of village assembly transcripts to examine gender and status-based patterns of influence in India and found a different role of women in public deliberations. Meanwhile, Mckenzie and Sansone (2017) compared performances of man and machine learning scores to entrepreneurs and their ability to predict entrepreneurship outcomes.

Furthermore, De Vitiis et al. (2016) have developed a methodology to assess, monitor and report FAO's statistics work by evaluating the use of agricultural data in evidence-based decision-making, as part of their Corporate Outcome Assessment.

The Statistical Literacy Construct from Watson and Callingham (2003) builds on the Structure of Observed Learning Outcomes (SOLO) taxonomy developed by Biggs and Collins (1982) to hierarchize statistical thinking into six stages of skills, which can be viewed as a progression of levels of statistical understanding. The strength of this model is that its statistical literacy scale has been widely validated by researchers, based on responses from a large number of students in Australia. At the top two levels of the Watson and Callingham (2003) construct, targets display abilities matching the critical thinking skills of the third tier of the Statistical Literacy Hierarchy in Biggs and Collins (1982). Klein et al. (2016) later adapted the original statistical literacy construct of Watson and Callingham to three consistent levels (non-critical, critical, and critical mathematical) and expanded the coverage to journalists.

The methodology developed within this paper adapted and improved the taxonomy in Klein et al. (2016) on newspapers in order to better measure the reference to and critical usage of statistical terms in national newspapers. The authors of this paper then expanded the dictionary of keywords for the first time and the methodology strengthened the statistical property of the final measurement. Moreover, this paper contributes to the existing literature by targeting policy documents.

2

2. METHODOLOGY

The methodology for the use of statistics in policymaking indicator seeks to reflect the complexity and breadth of policy documents. Part of the research agenda to measure data use includes, for example, focusing on new areas for developing and elaborating policies targeted at disadvantaged groups. For this, governments need to produce and thoroughly analyse the available evidence when designing policies. The new methodology relies on the fact that national development planning is now the rule, rather than the exception, in most countries. This allows improvements in policy formulation to be assessed over time.

The methodology was fully automated using a machine learning text-mining algorithm. The purpose of this was to systematically analyse multiple policy documents in a short amount of time. UoS is an internationally comparable indicator based on secondary sources for tracking progress towards evidence-informed policymaking, which can be updated on a regular basis.

A) KEYWORD SELECTION AND TEXT MINING

Automated text analysis, or text mining, is a technique for analysing a large corpus of documents in a systematic and efficient way. This paper uses the Wordscores method (Lowe, 2008), which scores documents based on the presence of a list of keywords defined from reference texts, to evaluate the use of statistics in policy documents. The scores present the probability of a document exhibiting a certain characteristic or fitting into a category (in this case, "evidence-informed policy"); therefore, the list of keywords affects the validity of the analysis. In the case of the UoS indicator, three corpuses were selected to define keywords: policy documents; international databases and frameworks containing monitoring indicators; and statistical dictionaries and coursebooks.

The limitations of text mining in comparison with semantic analysis are also discussed in the literature. The context in which a word or a sentence is placed can alter its meaning, and there is more than one single combination of words to convey a similar meaning. As the processes through which documents are produced are complex and no method can fully provide an accurate account, the results of any automated content analysis methods are imprecise but insightful (Grimmer and Stewart, 2013).

A clustering method was implemented to define a list of keywords from policy documents (Grimmer and Stewart, 2013). The algorithm, known as term frequency–inverse document frequency, is widely used in text mining practices. The method captures a list of words that are most frequently associated with different topics, such as a list of commonly used keywords in the corpus text, which are related to each topic. This was used for identifying "measurable concepts" that are present in these documents, as well as capturing how monitoring and evaluation arrangements are introduced and how the results of previous plans are discussed.

To improve the validity of the indicator, the preliminary list of keywords provided by the clustering has been complemented by a list of indicators from international agencies and a list of statistical terms related to the Sustainable Development Goals (see Annex 2). These databases contain data from national statistical systems and allow for comparisons over time and between countries. In addition, most countries have committed, for monitoring purposes, to international (regional or global) frameworks that include the indicators in their databases. A baseline list of statistical terms was identified from textbooks and dictionaries on statistics, such as the OECD Glossary of Statistical Terms (OECD, 2008).

Lemmatisation and stemming

Word lemmatisation (Plisson et al. 2004) was applied to all text and keyword lists and documents before proceeding with the analysis. Lemmatisation is the algorithmic process of determining the lemma of a word based on its intended meaning. Unlike stemming, lemmatisation depends on correctly identifying the intended part of speech and meaning of a word in a sentence, as well as within the larger context surrounding that sentence, such as neighbouring sentences or even an entire document.

Stop words

Stop words are usually defined as a set of commonly used words in any language, such as "the", "is", "at", etc. These words provide very little additional information for text analysis. By removing them during the pre-processing stage, the analysis can focus on the more important words. The stop words methodology is widely used in many applications.

The stop words list usually includes determiners (e.g., the, a, an), prepositions (e.g., above, across, before) and even some adjectives (e.g., good, nice). For different applications, the stop words list can vary greatly. For example, it is detrimental to remove too many stop words when analysing legal documents; a minimal stop words list is more suitable for that scenario. In sentimental analysis, the stop words list should be very carefully compiled with regard to adjectives. In other cases, including in our analysis, words such as "not" and "no" are not useful at all. Over the years, the text mining community has established a few widely used stop words lists. The text mining methodology in this paper uses a revised version of the Snowball stop word and stop words package of R.

B) SCORING SYSTEM

A "statistic" is considered to be any measurable value given in the report (i.e., absolute values, percentages, or fractions) at country level. It does not include exogenously determined variables (e.g., global interest rates), given geographical features (land area, coastline, etc.), or government financing figures (PARIS21, 2015: 3).

The use of statistics indicator consists of 4 sub-indices that aim to reflect the relevance of statistical evidence. It comprises four main dimensions: basic use (level 1), diagnoses (level 2), analytical use (level 3) and data disaggregation. Examples of identified sentences can be found in Annex 3.

- The sub-indices on three levels of use measure the extent to which statistics and statistical analysis have contributed directly to policy and decision making. The occurrence of a measurable term or keyword in the documents does not necessarily imply analysis (see Box 1 for an example). For the purposes of distinguishing the depth of the analysis of statistical evidence, the three levels were defined based on the Statistical Literacy Indicator (Klein, Galdin and Mohamedou, 2016), as follows:
- Basic consideration (Level 1) is the introduction of a measurable concept for indicating the state of and progress towards a specific development outcome (e.g., food production). Level 1 is then defined as a reference to key statistical indicators.
- Diagnosis (Level 2) is defined as the depiction of the state of such measurable concepts and the presumed causes for their variation (e.g., mortality rate of 5 per 1 000 adults, improvement of 5% in adult literacy). Level 2 refers to quoting actual numbers while referencing statistical concepts.
- Statistical Analysis (Level 3) is defined as using advanced mathematical concepts while referencing statistical concepts. A sentence is at Level 3 if it includes the use of statistical methods to depict the state of the measurable concept and the causes of its variation (e.g., the improvement of 5% in adult literacy rate correlates with improvements in life expectancy or the income of the lower quartile).

The methodology is based on a basic list of "measurable concepts" (see Annex 3) that is consistent throughout the three levels. For level 3, an additional list of "statistical terms" is added. When a keyword from the "measurable concepts" list is found together with one of the "statistical terms", the sentence is classified into level 3. If a "measurable concept" is found together with a figure (number), it is assigned to level 2. Finally, if a "measurable concept" is found on its own, it is assigned to level 1. An example of keywords found in a document can be found in Annex 3.

The three levels are mutually exclusive. To avoid double counting, texts are first evaluated at level 3. Then, sentences that are not identified as level 3 are evaluated at level 2, and finally at level 1. An overall score for the basic use of statistics is then obtained, being the sum of the three levels. In addition to the three levels, our analysis also searches for the use of disaggregation when a statistical term is referenced. The call to "Leave No One Behind" (LNOB), set out in the 2030 Agenda with its SDGs, has put forth an increasing need for better data and statistics to support informed policymaking. To the extent that data can make a difference to the way public policy is delivered, securing funding for statistics in the coming years is of the essence; good quality, disaggregated data is essential for putting LNOB in practice. Policy disaggregation was used as another dimension of the upstream measurement. A broader set of disaggregation keywords has been implemented to combine with the three levels, for producing a more subtle metric. The structure and keywords of disaggregation originate from the UNSD standards for data disaggregation.[3]

Texts were evaluated at sentence level, meaning that the scoring is based on the frequency of sentences that use keywords, instead of the frequency of keywords. Frequency counts of keywords were weighted proportional to the number of sentences in a document. Part-of-speech tagging was used to identify and distinguish past, present and future tenses.

The policy documents are scored by rating them following individual semantic analyses. For each sub-index in the table, scores are given based on the normalised relative frequency of keyword appearances in the document.

BOX 1: SEMANTIC DIFFERENCES BETWEEN TYPES OF SENTENCES

To measure the use of statistics in policy decision documents, it is important to differentiate the way in which statistical terms and indicators are used. Frequency in the use of a term is not necessarily related to use of statistics per se. As an example, the following two sentences were extracted from Zambia's National Development Plan 2017-2021 (Republic of Zambia, 2017: 21).

Sentence 1

"Therefore, the Government implemented a number of social protection interventions such as the social cash transfer, school feeding and the food security pack."

Sentence 2

"However, under the SNDP period a total of 111,400 low-capacity households received food security packs to support agricultural productivity for enhanced household food security, compared to 118,226 households under the FNDP period representing a drop of 5.8 percent involving 6,826 households."While sentence 1 only incorporates the concept of "food security" in the plan, sentence 2 provides a clear description of the indicator/statistic related to the programme

[3] *https://unstats.un.org/sdgs/files/Overview%20of%20Standards%20for%20Data%20Disaggregation.pdf*

C) NORMALISATION

For each sub-index in the table, scores are given based on the normalised relative frequency of keyword appearances in the document. Normalisation is required prior to any data aggregation as the indicators in a data set often have different units of measurement. A normal distribution indicates that the performance of policy documents with respect to the scorecard is somewhat even: there are few outliers, and the distribution is concentrated around the mean. A number of normalisation methods exist, the most common of which include ranking, standardisation (z-scores), max-min and categorical scale. Categorical scale is not suitable for the indicator because none of the indexes are categorical. The previous experience in the generic indicator showed that ranking countries proved to be less valuable than country self-diagnoses and self-comparison. Therefore, the analysis in this report focuses on two methods of normalisation: standardisation and max-min.

D) WEIGHTING AND AGGREGATION

Weighting can have a significant effect on the results of the indicator and the country rankings. A number of weighting techniques are recommended in this area by the OECD handbook. Some are derived from statistical models, such as factor analysis, data envelopment analysis and Unobserved Components Models (UCM), or from participatory methods like Budget Allocation Processes (BAP), Analytic Hierarchy Processes (AHP) and Conjoint Analysis (CA) (OECD, 2005). Regardless of which method is used, however, weights are essentially value judgements. While some might choose weights based only on statistical methods, others might reward (or punish) components that are deemed more (or less) influential, depending on expert opinion, to better reflect policy priorities or theoretical factors.

Most composite indicators rely on Equal Weighting (EW), i.e., all variables are given the same weight. This essentially implies that all variables are “worth” the same in the composite, but it could also disguise the absence of a statistical or an empirical basis, e.g., when there is insufficient knowledge of causal relationships or a lack of consensus on the alternative. In any case, equal weighting does not mean “no weights”, but implicitly implies that the weights are equal. Moreover, if variables are grouped into dimensions and those are further aggregated into the composite, then applying equal weighting to the variables may imply an unequal weighting of the dimension (the dimensions grouping the larger number of variables will have greater weight). This could result in an unbalanced structure in the composite index.

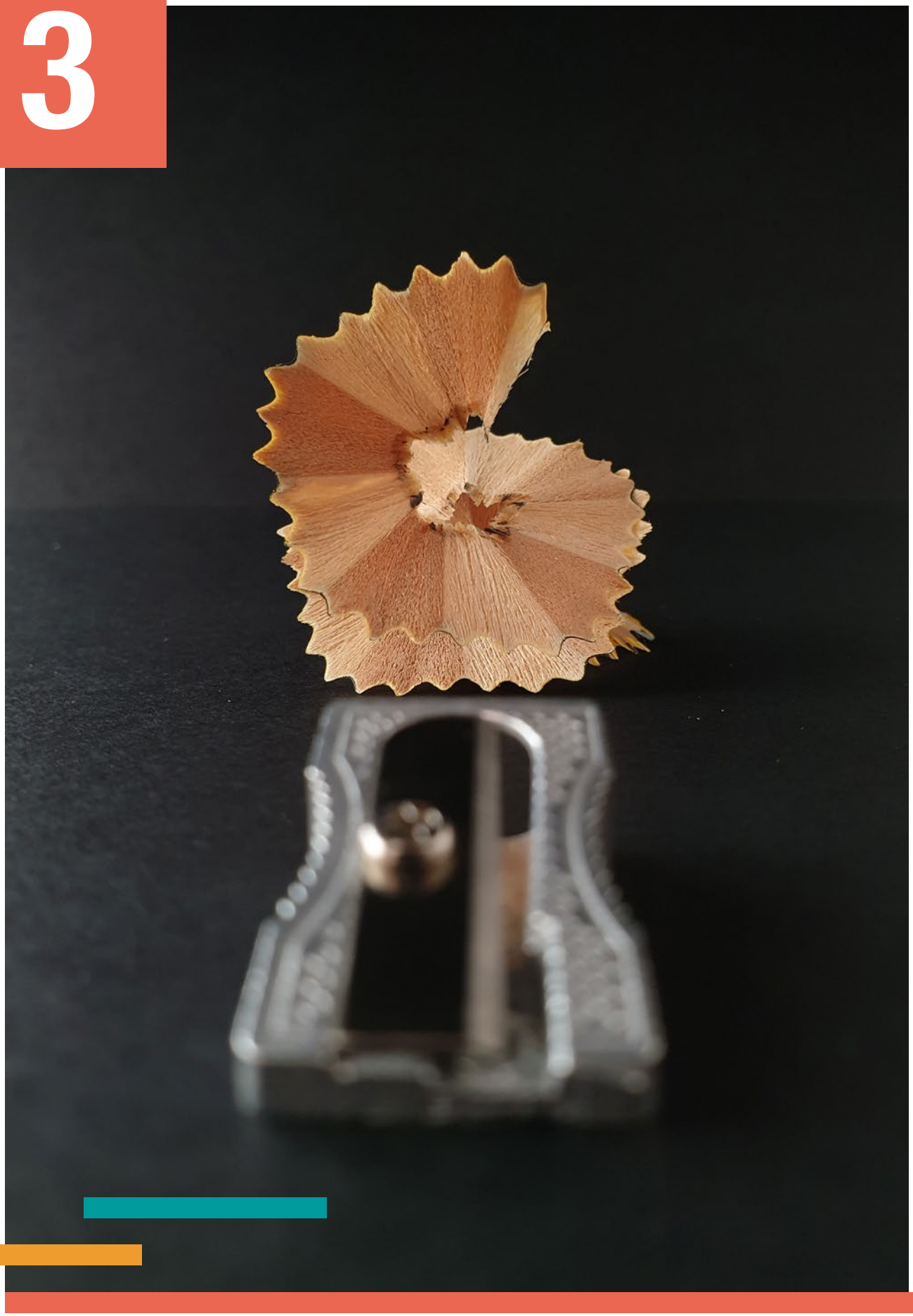
The tool developed by PARIS21 uses Principal Component Analysis (PCA) for its weighting methodology, given its statistical quality over other models. A non-weighted score and weighted score chosen by the experts are also provided for comparison. Based on the perspective of different sub-sectors, users can also choose to use customised weights for the indicator. If a user decides to choose customised weights, there is a risk of, in addition to low statistical quality, not reflecting how governments approach the use of statistical evidence in policymaking, but instead creating a theoretical construct which is detached from “reality”.

Aggregation methods also vary. The most commonly used method, the linear aggregation method, is useful when all individual indicators have the same measurement unit, provided that some mathematical properties are respected. Geometric aggregations are better suited if the indicator needs to reflect some degree of non-compensability between individual indicators or dimensions. While linear aggregations reward base indicators proportionally to the weights, geometric aggregations reward those countries with higher scores. An undesirable feature of additive aggregations is the implied full compensability, such that poor performance in some indicators can be compensated for by sufficiently high values in other indicators. In our case, different levels of use and disaggregation are better captured by geometric aggregation due to the non-compensability of the data.

Further methods have also been implemented to improve the reliability of the index on the use of statistics. We use certified methods to validate and check robustness of our indicator, following the OECD/JRC Handbook on Composite Indicators. A cluster analysis with criteria such as national GDP, regions, and Official Development Assistance (ODA received) were applied to test the validity of the indicator. As one of the most used methodologies on weighting composite indicators, Cronbach Alpha was implemented, and the Monte Carlo Method (MCM) was also used by randomising all assumptions taken on the index simultaneously.

Previous versions of the UoS indicator included two additional dimensions in the composite indicator, calculated manually: a) a dichotomic sub-index, reflecting if the national policy document contained a Monitoring & Evaluation section, and b) a dichotomic sub-index, indicating if the document referred to previous editions. For the purposes of the current analysis, provided these two dimensions were not considered indispensable to all policy documents, and were not measurable through an automatised process, they were excluded. Future versions of the indicator could include this type of component to reflect the quality of policy documents.

3



3. RESULTS

DESCRIPTIVE STATISTICS

The text mining analysis was applied to 203 PRSPs and NDPs belonging to 102 countries, which were written in English (Table 2 provides a more detailed distribution) and publicly available.

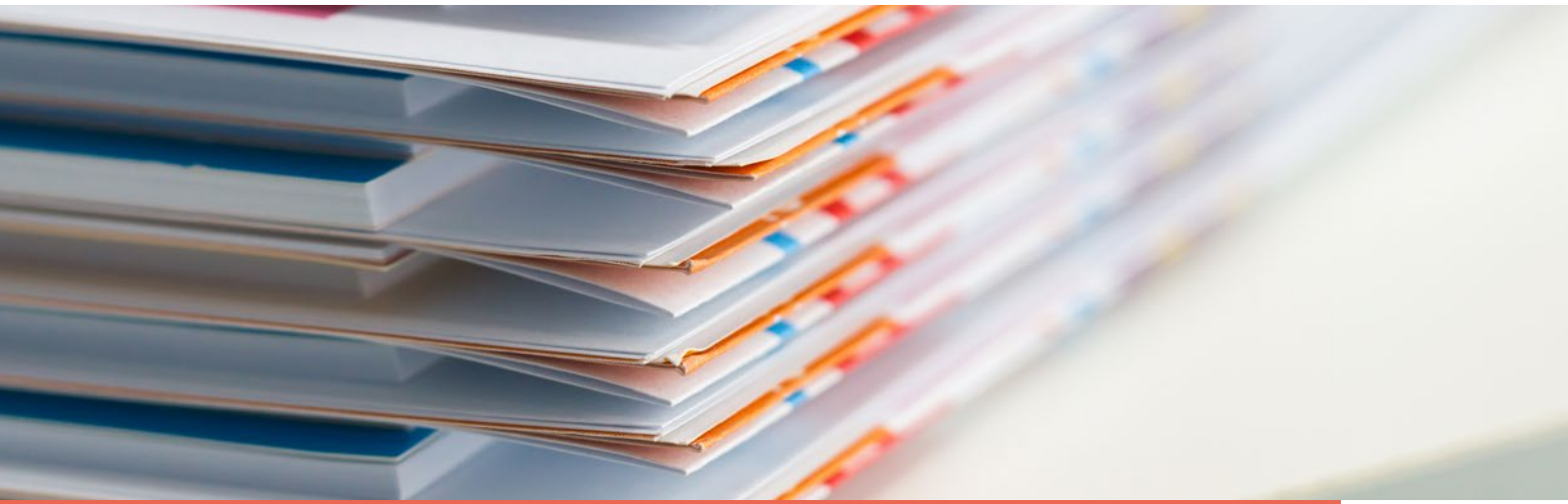
Table 1: Countries and documents per continent

Region	Countries	Documents
Asia	31	71
Africa	44	97
Americas	11	15
Europe	7	9
Oceania	7	11

The period covered by the indicator spans from the year 2001, the beginning of the Highly Indebted Country Initiative (HICI), to 2020. Because countries follow different timetables for sanctioning these plans, two waves were devised: the first between 2001 and 2010 and the second between 2011 and 2020. Table 3 provides more detail on the number of documents per wave and regional distribution.

Table 2: Distribution of countries by wave and region

Region	Wave 1	Wave 2
Asia	40	31
Africa	52	45
Americas	8	7
Europe	6	3
Oceania	7	4



In total, the sample includes policy documents for 100 countries distributed in two waves in the corpus of documents. Figures 1.a and 1.b present the use of statistics scores for the second wave and the difference between the average scores for waves 1 and 2 by country.

Figure 1.a: UoS scores in the policymaking indicator for plans published from 2009-2017

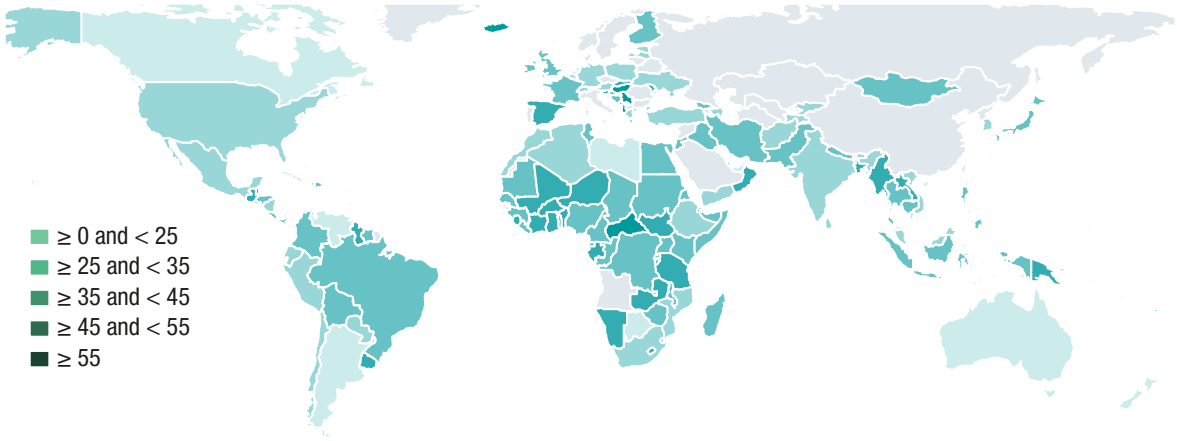
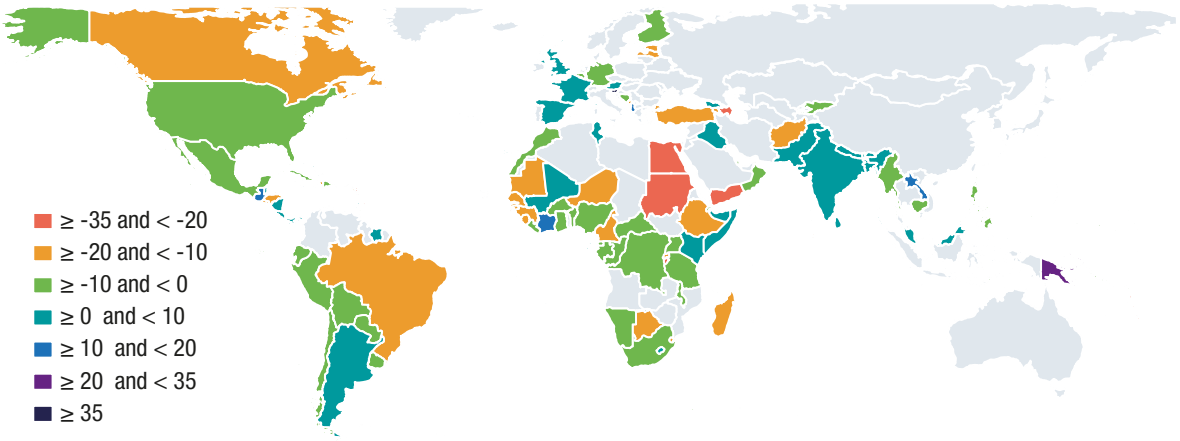
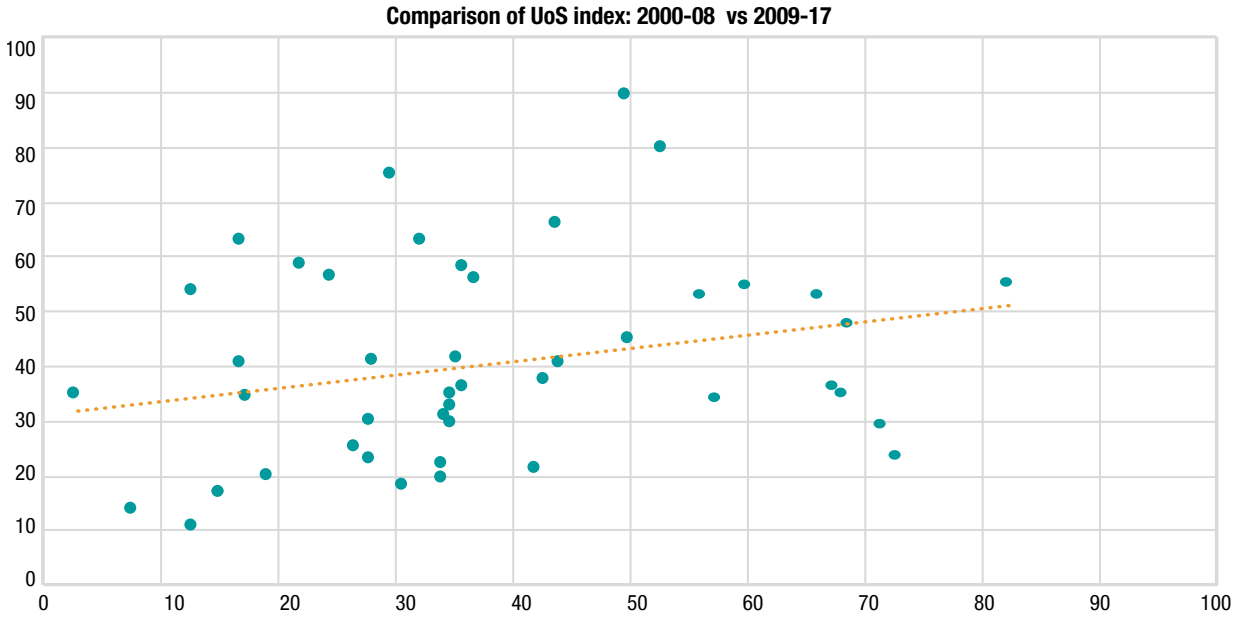


Figure 1.b: Change in UoS scores in the policymaking indicator between plans published from 2000-2008 and plans published from 2009-2017



The correlation between country scores in the two waves was relatively low (Pearson's r-coefficient = 0.25, Figure 2), suggesting a non-linear progression of the use of statistical evidence for policy planning. Other factors linked to countries' policy priorities, governance or political dynamics could explain this result, which will be the subject of further exploration.

Figure 2: Scatterplot and fit line between scores in wave 1 and wave 2



The mean score of the UoS indicator in the current sample is 39.05 points, with a standard deviation of 18 points. It is important to notice the selected aggregation method for the composite indicator is an arithmetic sum of the components, which could affect the final score for countries. As a robustness check, different aggregation methodologies and weighting methods are implemented in the following section.

The minimum score is 2.8 and the maximum is 88.9. Considering the regional distribution of the score, the scores for Africa, Americas and Oceania are similar (41 points), followed by Asia (35 points) and Europe (29 points). These differences may be explained by the fact that some European countries do not have NDPs or their NDPs focus on qualitative, strategic documentation.

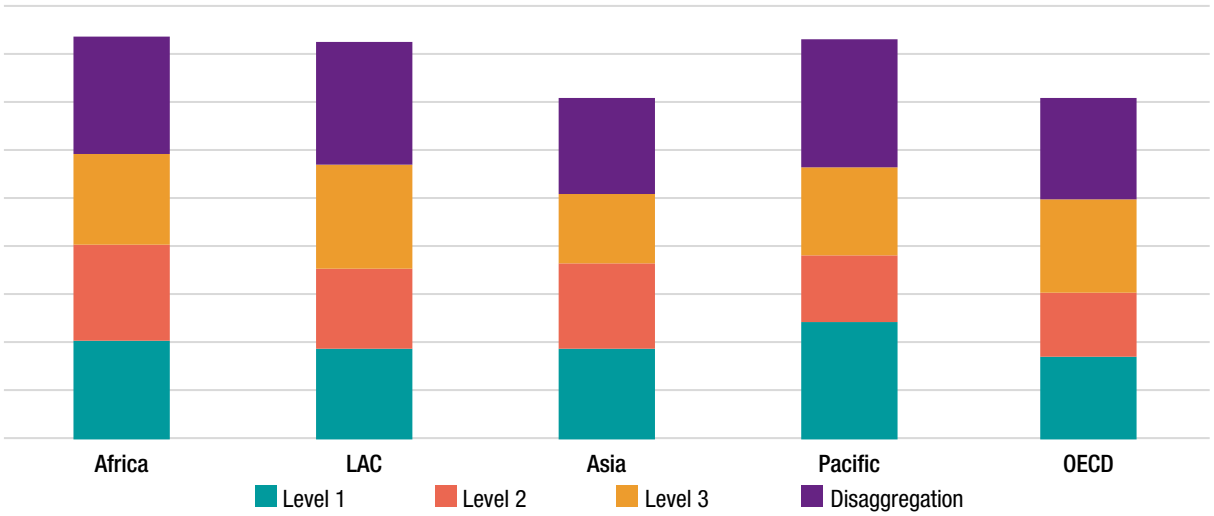
Table 3: Average UoS score by region and wave

Region	PRSPs		NDPs		Total	
	Wave 1	Wave 2	Wave 1	Wave 2	Wave 1	Wave 2
Asia	37	37.2	30	36.5	35.1	36.6
Africa	43	39.6	49.8	39	43.7	39.4
Americas	53	16.1	51.2	33.1	52.4	28.2
Europe	23.1		35.5	32.9	27.3	32.9
Oceania	33	47.1	36	55.2	35.2	51.1

For most regions, Level 1 data use and disaggregation represents the largest share of the indicator, with most regions achieving the highest scores in these two categories. Level 3 and Level 2 data use, which denotes the more complex use of data for statistical inference, represents a smaller share of the overall index. Interestingly, the disaggregation component of the index, which measures how national policy documents track disaggregation-related concepts such as such income or gender, is similar across policy documents for most regions. This does not fully reflect the availability of disaggregated data across regions, raising questions on disaggregated data use for policymaking. Over time, there has been an improvement overall in the use of statistics in national policy documents.

The figure below illustrates the aggregate indicator scores for the second period. The significant difference (at 5% level) between periods in the average country score suggests that most countries have improved their use of statistics in national planning and policymaking since 2000.

Figure 3: Distribution of scores by region

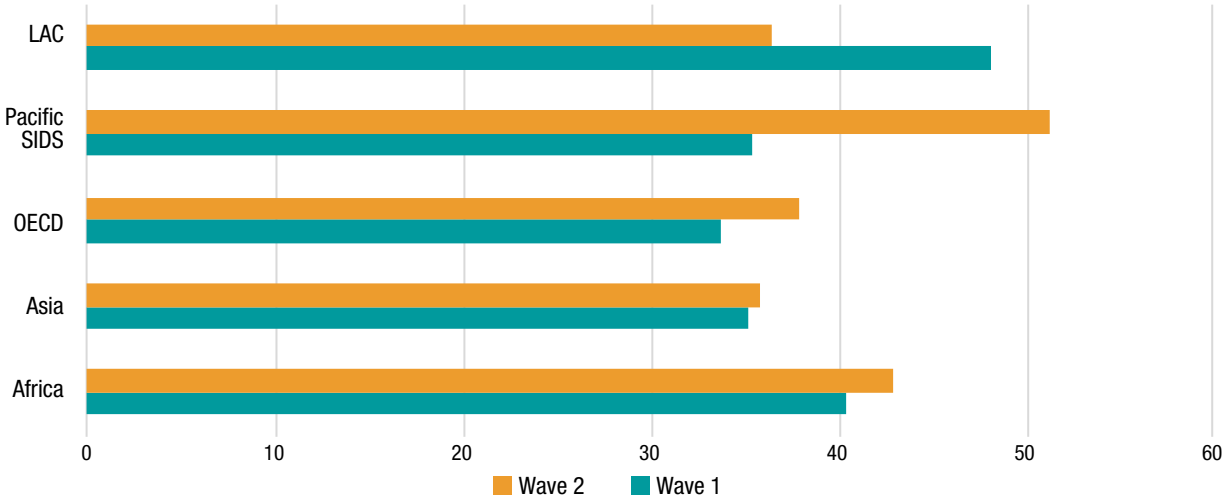


The regional averages for the indicator between the 2 periods (Table 2b) show that, a priori, there are relatively few improvements in the index for most regions, with the exception of Oceania. These modest improvements are somewhat surprising, as it is expected that certain countries will have more elaborate and data rich NDPs and PRSPs with time. The difference in score between PRSPs (37.9) and NDPs (39.8) does not indicate large differences in the way these two different reports use data and statistics. Considering the two waves, differences in average values for the index are also small (38.3 for PRSP and 40.3 for NDP in wave 1, 37.7 for PRSP and 38.2 for NDP in wave 2). Table 4 depicts the differences between groups. The methodology was also proved to be expandable to sectoral statistics and for different disaggregations (Box 2).

Table 4: Summary of statistics by continent, wave and type of document

	Minimum	Maximum	Mean	Standard deviation
Full sample	2.8	88.9	39	18.8
Asia	5.9	81.9	35.8	17.9
Africa	2.8	88.9	41.7	19.2
Americas	6.6	79.1	41.1	19
Oceania	14.2	74.6	41	22
Europe	18.3	53	29.1	11.4
2000-2008	2.8	82.4	39.9	19.2
2009-2017	4.0	88.9	37.9	18.3
PRSP	2.8	88.9	39.7	19.3
NDP	5.9	79.3	37.9	18.1

Figure 4: Comparison of scores in the 1st and 2nd periods



The results also suggest that data use is not always associated with economic or statistical development. The performance in terms of referring to and critically using statistical terms is not associated with basic economic indicators, governance indicators or other indicators of statistical development. While the indicator can provide an estimation of the quality of a national policy document, it cannot provide a detailed account of how data drives policy decision making. The next section will present more results on this.

BOX 2: MEASURING USE OF GENDER STATISTICS IN POLICY DOCUMENTS

Policies can affect different groups of people disproportionately, directly or indirectly. Gender inequality serves as a key example. If policy documents are written as gender-neutral, opportunities will be missed to respond to differing needs, experiences and outcomes for men and women in national development. Moreover, policies should adopt an intersectional lens by including the views of different sub-groups of men and women in the formation and delivery of those policies. Without this, governments risk misidentifying the different effects of policy on various subpopulations and the systems and organisations that support them. References to gender equality and gender statistics in a policy document can ensure the gender dimension of policies is explicitly adopted, described and evidence based.

In collaboration with UN Women, PARIS21 expanded the methodology on measuring the use of statistics in policy documents to include gender statistics. In order to ¬¬ the scope of this exercise, it is important to mention its definition as developed by the UNSD, since gender statistics are more than data disaggregated by sex. Gender statistics are defined by the sum of the following characteristics:

a. Data are collected and presented by sex as a primary and overall classification;

b. Data reflect gender issues;

c. Data are based on concepts and definitions that adequately reflect the diversity of men and women and capture all aspects of their lives;

d. Data collection methods take into account stereotypes and social and cultural factors that may induce gender bias in the data (UNSD, 2016).

Specialised keyword lists were redeveloped to make it more suitable for re-evaluating the dimension of gender, taking into account possible implications of these characteristics in policy documents. PARIS21 presented the updated methodology and keyword lists to a group of experts in October 2019. Based on the comments and input, the final modified keywords lists were validated in December 2019.

The preliminary results show that in the 55 countries analysed, the use of gender statistics increased slightly in the past decade (2011-2020). Americas and Europe lead in overall scores, while Asian countries have a higher average score in quoting actual numbers when referring to gender statistics. The increase in the use of gender statistics in Asian countries significantly correlated with the increase in their Government Effectiveness scores in the World Governance Index.

The correlation between country scores in the two waves was relatively low, suggesting a non-linear progression of the use of statistical evidence for policy planning. Other factors such the adoption of the SDG framework, the recent trend of the data revolution, the changes in national policy priorities, and governance or political dynamics could explain this result, which will be the subject of further exploration. Further discussion on these findings is available in Annex 1 of this document.

ECONOMETRIC ANALYSIS

Having estimated the UoS indicator, we test whether the indicator correlates with other variables of interest. Four main factors related to policy formulation are mentioned in the literature (ODC, 2015): a) political context and institutions; b) credibility of the evidence; c) links between policymakers with other actors; and d) external influences (such as culture, international relations, economic factors, etc.) (Crewe and Young, 2003).[1] Table 5 depicts the correlation between selected indicators that represent those factors and the UoS indicator. Some of the displayed correlations are analysed in the remainder of this section.

Table 5: Pairwise correlation (r) between selected indicators and use of statistics for the various sub-populations in the sample

	SCI (Yr-1)	CPIA (Yr-1)	WGI: Voice and Accountability (Yr-1)	Press Freedom (Yr-1)	HDI (Yr-1)	Literacy (Yr-1)	GDP per Capita (Yr-1)	Official Development Assistance (10 prev. yrs)
Total	-0.1*		0.05		-0.17*		-0.17*	-0.13
Africa	0.17		0.19		0.10		0.00	0.04
Asia	-0.37		-0.04		-0.34		-0.25*	-0.22
Pacific	-0.28		-0.27		-0.22		-0.07	0.49
OECD	0.43		0.59		0.42		0.53	0.05
LAC			-0.13		-0.27		-0.38	-0.19
PRSP	-0.09		0.07		-0.18*		-0.13	0.00
NDP	-0.10		0.05		-0.19		-0.25*	-0.25*
2000-2008	0.20*		0.09		-0.06		0.01	-0.19*
2009-2017	0.09		-0.01		-0.31*		-0.29*	0.00

* 5% significance

Note: SCI (2017) is the World Bank's Statistical Capacity Indicator (total) for the year prior to the plan being sanctioned. CPIA is the World Bank's Country Policy and Institutional Assessment (2017) for the latest available year prior to the sanctioning of the plan. WGI is the World Bank's World Governance Indicators (2017) for the year prior to the sanctioning of the plan. Press Freedom is the Freedom of the Press score from Freedom House (2017) for the year prior to the sanctioning of the plan. HDI is the Human Development Index from UNDP (2016) for the year prior to the sanctioning of the plan. Literacy is adult literacy rate from UNESCO's UIS database (2017). GDP per Capita is in PPP for the year prior to the sanctioning of the plan, from the World Bank's World Development Indicators (2017). ODA is the sum of gross official development assistance in PPP for the 10 years prior to the sanctioning of the plan from OECD.Stat (2017).

The results (Table 5) suggest the UoS score can be mainly linked to the quality of the statistical evidence of institutions, and to external "cultural" influences. The Statistical Capacity Indicator (SCI) of the World Bank provides a proxy for quality of statistical evidence. It assesses the ability of a country to collect, analyse and disseminate high quality data, assigning scores that range from 0 to 100. There is a statistically negative correlation between the UoS indicator and the SCI. This result could be counterintuitive, as in principle, a country's capacity to produce statistical evidence should be accompanied by more frequent UoS in official documents. There is a statistically significant (at 5%) positive correlation with the UoS indicator for the period 2000-08. This relationship can be observed individually with the three subcomponents of the SCI (source data, methodology and periodicity).

In terms of the institutional environment, for African countries, the score is also positively correlated (r= 0.27, significant at 5%) with the average rating of the Country Policy and Institutional Assessment of

[1] *More information on the initiative can be found on: https://www.odi.org/our-work/programmes/research-and-policy-development*

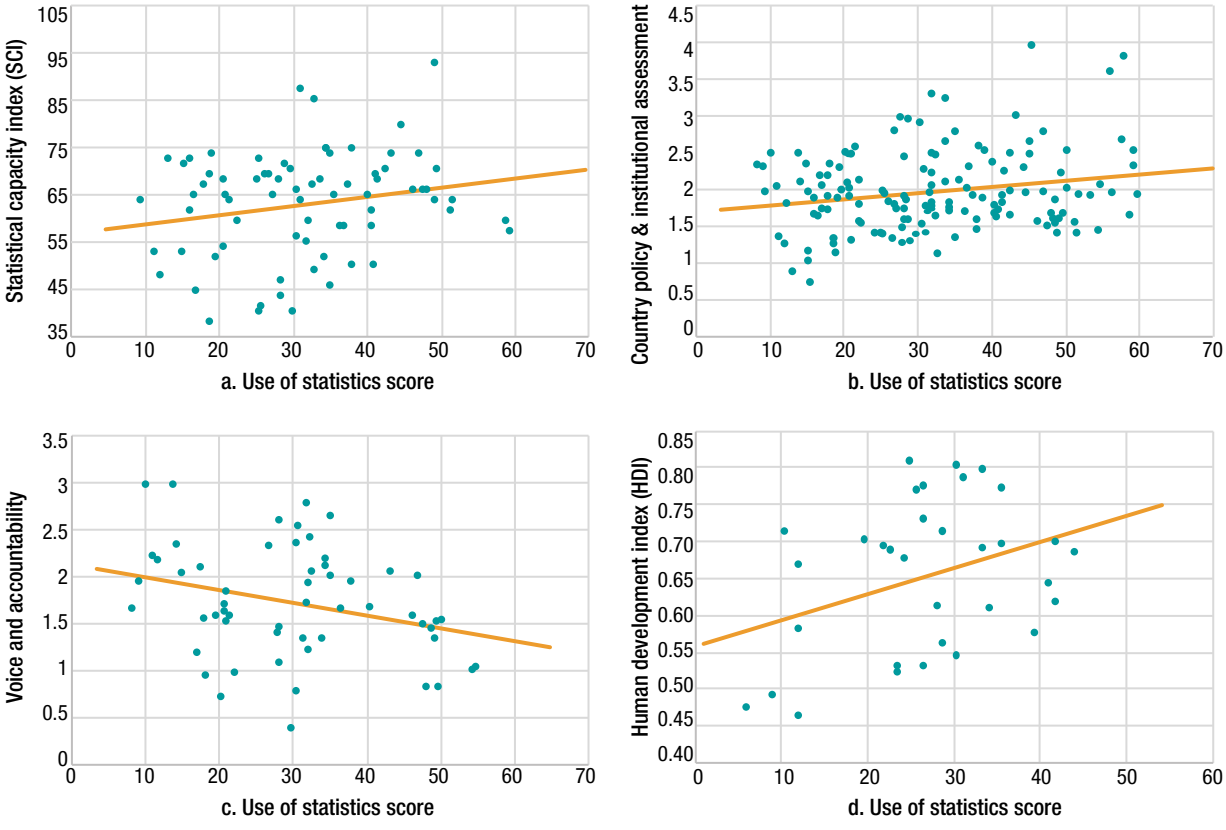
the World Bank. The assessment rates countries from 1 to 6 in terms of their economic management, structural policies, policies for social inclusion and equity, and public sector management and institutions. Indeed, the length of appointment of the chief executive negatively correlates with the use of statistics in policies (r=-0.20, significant at 10%). This effect disappears when those who have been in power for more than 20 years are excluded.

In Asia, the institutional environment is also relevant, yet with an inverse effect. A higher use of statistical evidence in national planning is related to less freedom of civil society. The total score of UoS correlated slightly (r= 0.05) with the "Voice and Accountability" component of the World Governance Indicators of the World Bank (Table 5). This indicator measures the extent to which citizens are able to select their government and their civil liberties are respected. This relationship can also be observed when comparing the Freedom of the Press Report from Freedom House. The indicator scores countries from 0 (highest) to 100 (lowest) on media freedom. The correlation is of r= 0.33 is also significant at 1%. While this finding would require further exploration, this paper focuses mainly on testing whether the UoS indicator correlates with established measures of institutional quality.

Finally, the Human Development Index (HDI) of UNDP significantly correlates with the UoS in policymaking in Oceania, Europe and America (r= -0.17, 5% significance). The HDI includes three components: income (Gross National Income per capita), education (years of schooling) and health (life expectancy at birth). In particular, adult literacy (UNESCO) strongly correlates (r= 0.63) with the use of statistics in policymaking (significant at 1%) for these three continents.

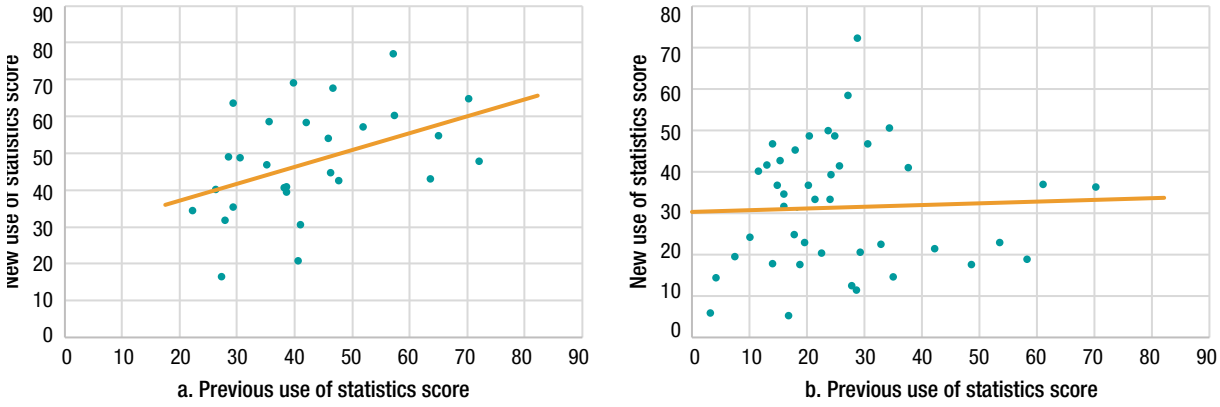
In sum, the UoS indicator does not seem to be linked to basic economic indicators (e.g., GDP), governance indicators (e.g., CPIA index) or indicators of statistical development (e.g., Statistical Capacity Index). Correlation with the HDI index is negative and significant for most regions and waves.

Figure 5: Scatterplots with linear fit between UoS and relevant indicators (both waves)



To test whether improvements in the UoS score depend on political factors, a comparison was made between the UoS index for national plans written under the same executive head (e.g., prime minister, president, chancellor) and those written under different heads. [2] Figure 6.a shows a correlation between the UoS score in a given year, and the average score of previous plans under the same head (r= 0.23, significant at 10%). This effect disappears with a change in the executive (r= 0.08, not significant, Figure 6.b). This could suggest that institutional knowledge on using evidence for policy planning is not automatically transferred between administrations, which could also explain why there is no correlation between the scores of the two waves (Figure 1).

Figure 6: Comparison of UoS total score in a given year with the average of previous scores under the same head of the executive power (a) and between diffferent heads (b)



[2] *The Inter-American Development Bank's Database of Political Institutions (2017) records several variables related to the executive and legislative power of countries. One of these is the number of consecutive years the head of the executive power has been in office. If the head has been in office for the same or a longer amount of time than the time span between two plans (i.e., the time between the sanctioning of one plan and the following one), it is assumed that both were written under his/her mandate. Otherwise, it is assumed that both plans were written under different heads.*

4

4. ROBUSTNESS CHECKS

This section revisits the results for the UoS index and introduces robustness checks, with the aim of providing insight for future improvements. Two types of analysis are presented: a) construct validity checks (how well the methodology reflects the theory behind the indicator and b) an uncertainty and robustness analysis for composite indexes, following the methodology proposed in the OECD Handbook on Constructing Composite Indicators (2005) and the European Commission's Tools for Composite Indicators Building (2005). Preliminary results suggest that the indicator is valid (it is in line with its purpose) and sensitive to the weights, components and aggregation methods.

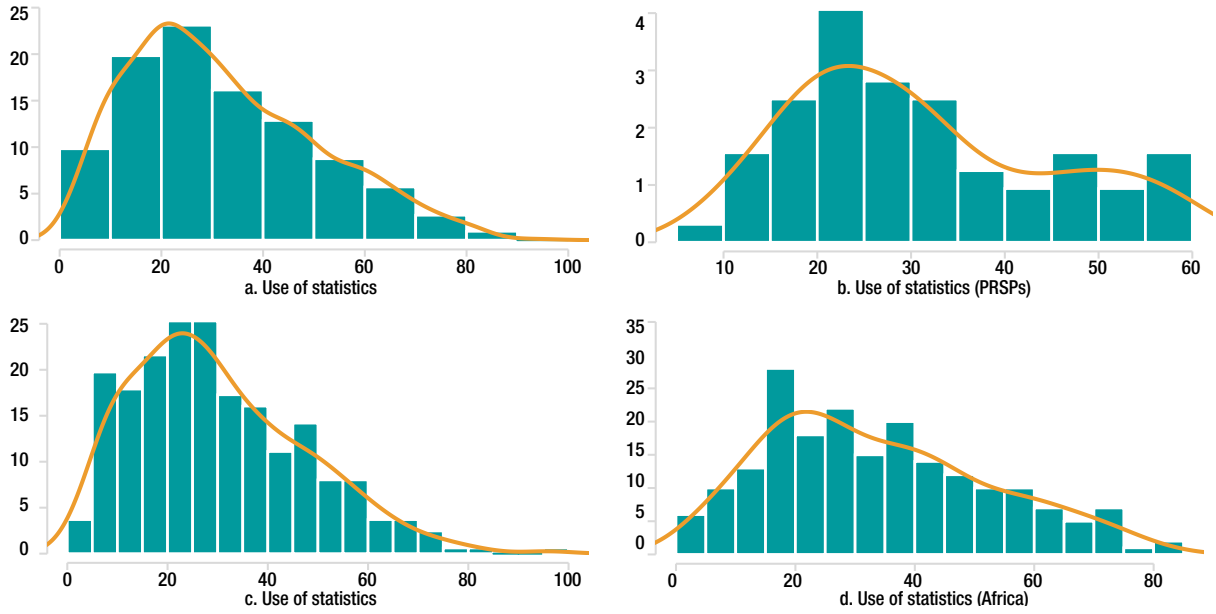
A) CONSTRUCT VALIDITY

In order to test whether the indicator measures the use of statistical evidence in policymaking, the distribution of the indicator is analysed in the first instance. The policy documents are scored with a pre-defined scorecard (content analysis), rather than rating them following individual semantic analyses. Because of this approach, there is a risk of not reflecting how governments approach the use of statistical evidence in policymaking, instead creating a theoretical construct that is detached from "reality".

The distribution of total scores including all the components that make up the indicator, as depicted in Figure 2 (a), is skewed left (Shapiro Wilk test p= 0.00 indicates that the distribution is not normally distributed). Beyond its statistical properties, a normal distribution would indicate that the performance of policy documents with respect to the scorecard is somewhat even: there are few outliers, and the distribution is concentrated around the mean.

A closer inspection reveals that there are subpopulations within the sample. While the scores of the PRSPs (3.b) are not normally distributed (p= 0.032), neither is the case for NDPs (p= 0.01). The scores of each individual continent are normally distributed, as illustrated by those of Africa (3.c) and Asia (3.d), with the exception of Europe (p= 0.02). None of the waves are normally distributed (wave 1, p= 0.01 and wave 2, p= 0.02).

Figure 7: Histogram and kernel density of the aggregate of the three levels and the total score



Note: These results show both waves combined, for illustrative purposes.

A further check to validate the methodology is conducted by evaluating correlations between individual components of the indicator, and correlations between the indicator with key economic and statistical variables. The aim is to determine whether there is any double counting that could bias the total results, following the Handbook on Constructing Composite Indicators (OECD, 2005). Table 5 shows the correlation between individual components of the UoS indicator. Although there is no suggested threshold to determine whether two components are highly correlated, only levels 2 and 3 have a high enough r coefficient to require further exploration.

Table 6: Pairwise correlation coefficient (r) between the components of the indicator

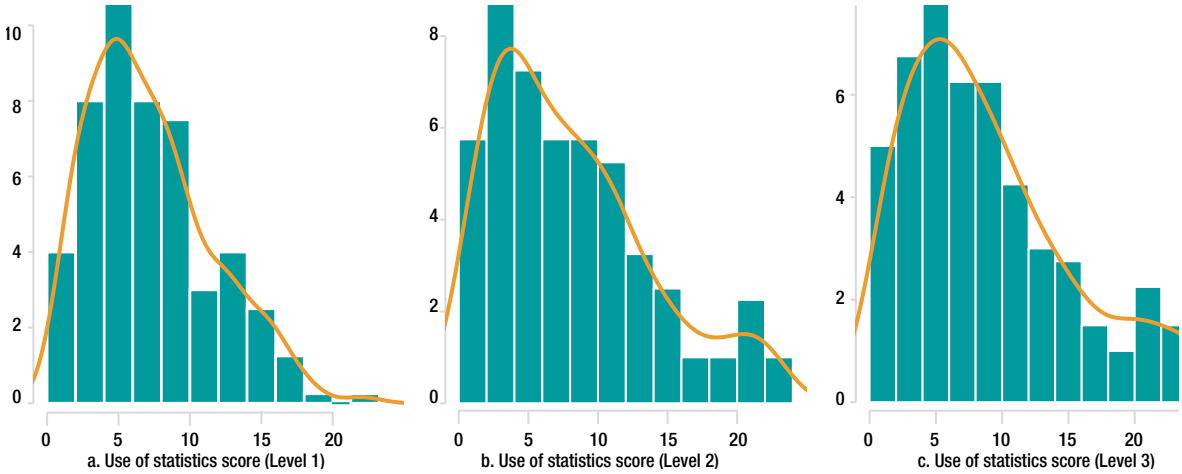
	Level 1	Level 2	Level 3	Total
Level 2	0.67*			
Level 3	0.75*	0.50*		
Disaggregatio	0.92*	0.57*	0.84	0.94

*5% significance

It is worth noting that the three levels constitute the core of the UoS indicator: they signal to what extent the decisions leading to the NDP or PRSP were informed by statistical analysis. A greater reference to statistics in the text denotes that evidence plays a more relevant role at the time of decision making. In this sense, it is desirable for the total score of the indicator to be largely influenced by the aggregate of the three levels.

Each of the three-level sub-indices provides additional information for the indicator. While Level 1 captures the simple use of a measurable concept (e.g., literacy), level 2 captures the use of a keyword with an actual value and the third level captures the use of a keyword with a statistical concept (e.g., correlation). A high score in a single level does not provide an adequate picture of the extent to which available data was used. None of the levels, when taken individually, have a distribution that resembles their aggregated score (none of the levels have a normal distribution, $p<0.05$). The three sub-indexes complement each other and should be considered together.

Figure 8: Histogram and kernel density of the individual levels



Note: Estimation for full sample (199 countries) in waves 1 and 2.

The total score strongly correlates with the each of the three levels (0.94, 0.76 and 0.88, respectively). The average score for the three levels together is 27 with a standard deviation of 13. The minimum score is 2.8 and the maximum is 63. The slight differences between the total score and the aggregate score by levels is explained by the scores in the disaggregation component.

B) ROBUSTNESS CHECKS

Considering the composite nature of the UoS indicator, alternative designs should be tested for verifying the robustness of the scoring method. The change in countries' ranking (in this case, of policies) is the output of interest. For this, we follow EU's Tools for Composite Indicators Building (Nardo et al., 2005) to assess potential transformations of the UoS indicator, including normalising scores, excluding components and assigning different weights.

All composite indicators contain a degree of partiality; however, they are useful for guiding policymakers and practitioners. Poor construction would send a misleading policy message, but such a case could occur if, for example, experts disagreed on the relevance of each component (not only in terms of weight, but of inclusion) or if the aggregation method changed in subsequent editions. The interpretation of the scores and ranks is constrained by the assumptions that are made at the time of their construction. For this reason, consensus is a precondition for UoS to make a meaningful contribution to supporting the practice of evidence-informed decision-making.

The following transformations could be beneficial in order to investigate the sources of uncertainty that could be applied (although not all of them are relevant to the UoS indicator): inclusion and exclusion of sub indicators – components, in our case – modelling of data error, alternative editing schemes, alternative data normalisation schemes (such as rescaling, standardisation, use of raw data), weighting schemes and aggregation systems (Nardo et al., 2005: 86).

The more transformations the indicator suffers (in terms of exclusion-inclusion of components and changes in weights and aggregation), the larger the difference between the original relative position (rank) and the new one. As shown in Figure 7, the lowest ranks (lowest scores) suffer less variation than the highest ones, but both are more stable than those in the middle. This phenomenon has been observed in other composite indicators (Cherchye et al., 2008). For example, Nigeria's 2004 PRSP moves to rank 141 from rank 92 when level 2 is eliminated. When comparing alternative specifications in terms of the variation in individual ranks (uncertainty analysis), the effect is more marked. Table 6 shows how the difference ranks with respect to the original specification of the indicator. The exclusion of levels 2 and 3 cause the largest variation, taking into account that a geometric aggregation is not possible for all observations. As shown in the scatterplots, the highest or lowest ranks remain, while those around the middle positions are greatly affected. From the selected policies, Albania's 2001 PRSP, originally ranked 125, suffers an average absolute shift of 29 positions.

Table 7: Shifts in rank for alternative specifications of the indicator

Rank	Country	Year	Diff Rank 2	Diff Rank 3	Diff Rank 4	Diff Rank 5	Diff Rank 6	Diff Rank 7	Diff Rank 8	Diff Rank 9	Diff Rank 10	Avg diff.
2	PHL	2017	-2	-3	-5	-2	-1	-4	-1	-4	-1	2
12	CPV	2004	+4	-16	-5	-12	+11	+10	-44	-47	-13	18
22	LAO	2016	-19	+6	0	+7	-47	-49	14	+13	+9	18
36	BGD	2011	+1	-19	-9	+17	-24	-29	-14	-16	-10	15
47	BTN	2013	+12	-38	-4	+16	-6	-9	-32	-37	-26	20
66	NAM	2012	+13	-36	-13	+27	-9	-18	-21	-27	-23	20
86	ARM	2001	-54	+42	+16		-23	-10	+21	+37	+32	29
92	NGA	2004	-2	-4	-4		+20	+19	-49	-52	-7	20
125	ALB	2001	-6	+26	+38	+74	-39	-20	+3	+34	+17	29
141	PAK	2001	-11	+21	+8		+22	+30	-19	-20	+9	17
159	ERI	2004	-9	-4	+1	+94	-7	-2	-13	-11	-1	16
197	PNG	2004	0	+1	0		+1	+1	+2	+1	0	1
Average difference			8	11	6	30	10	10	12	13	7	12

Rank 2: Non-weighted aggregate

Rank 3: Weighted aggregate of standardised scores (z scores)

Rank 4: Non-weighted aggregate of z scores

Rank 5: Weighted geometric aggregate (only for documents where all components are different than zero)

Rank 6: Aggregate excluding level 3, rescaled weights for the rest of the components

Rank 7: Aggregate of z scores, excluding level 3 and rescaling weights for the rest of the components

Rank 8: Aggregate excluding level 2, rescaled weights for the rest of the components

Rank 9: Aggregate of z scores, excluding level 2 and rescaling weights for the rest of the components

Rank 10: Weighted aggregate of rescaled scores for individual components $I_{q,c} = \frac{X_{q,c} - min(X_q)}{range\ (X_q)}$

A comparison of the original scoring method and alternative specifications confirms that, when approaching the middle of the distribution (the first column of Table 6), policy documents' ranks suffer the greatest variation with alternative specifications of the indicator. When comparing the variation between the extreme (top and bottom) and middle positions, the authors found the first six policy documents remain within the top six and the bottom seven remain at the bottom, while those ranked between 40 and 50 in the original ranking shift by as many as 37 positions (Armenia's 2001 PRSP).

The relationship between the alternative specifications of the UoS indicator and these factors was tested, and no perceivable differences were found. This indicates that, even with variation in ranks, the performance with respect to relevant factors affecting the uptake of evidence in the policymaking process is consistent.

5

5. CONCLUSION

"Fully integrating statistics in policymaking" is a common objective for actors in the national statistical system and policymakers in governments. The UoS indicator, first introduced by PARIS21, has gained relevance in recent policy discussions as an instrument to monitor progress towards evidence-informed policymaking. This document presents the methodology and innovations to the indicator with the aim of tracking key dimensions for monitoring the Cape Town Global Action Plan.

Based on a comprehensive text mining technique, the indicator screens a large set of NDPs and PRSPs to provide insights on the role of data and statistics in policy formulation and policy planning. Countries envision NDPs as a roadmap for achieving large positive impacts on wellbeing. Providing guidance on how to use solid statistical evidence for the allocation of public resources to specific policies, and tracking progress towards development objectives, is essential for improving their chances of succeeding.

An important contribution of this work is to improve the way in which the indicator measures the concept of use of statistics. Three levels of "basic use" were envisioned to differentiate the occurrence of a keyword from the analysis of statistical evidence. Building on the existing literature on Statistical Literacy, a similar methodology was incorporated to analyse national development plans through machine learning techniques. Other methodological improvements included the introduction of a "clustering method" to ensure that statistical keywords used in the data mining process correspond to the public documents in the sample.

The results introduce a robust indicator that is relatively robust to different weights and aggregation methods. The indicator was designed to capture both the use and the complexity of statistical evidence in the definition of policies. It also explores the important dimension of data disaggregation in national development plans, one of the key areas for the implementation of the 2030 Agenda.

The findings also suggest that country scorings are not necessarily permanent over time, which could suggest different appraisals of statistical evidence across time or lack of knowledge transfer between different administrations. Regional performances vary, as do the factors associated with the use of statistics.

The UoS indicator provides a more comprehensive analysis, including key dimensions for the implementation of the CT-GAP. In the future, the indicator will develop specific modules, including monitoring and evaluation frameworks, impact evaluation and assessing progress across time. Further work will also include developing indicators for specific sectors (e.g., agriculture, gender or labour), using sector-specific lists and documents for the analysis.

REFERENCES

Ardanaz, M., Scartascini, C., and Tommasi, M. (2010), "Political institutions, policymaking, and economic policy in Latin America", Inter-American Development Bank *Working paper* No. IDB-WP-158, http://www20.iadb.org/intal/catalogo/PE/2010/04914.pdf.

Cherchye, L. , Moesen, W., Rogge, N., Van Puyenbroeck, T., Saisana, M., Saltelli, Liska, R., Tarantola, S. (2008), "Creating composite indicators with DEA and robustness analysis: The case of the technology achievement index" *Journal of the Operational Research Society* 59 (2): 239 – 251

De Vitiis, C., Donmez, A., Dowlatchahi, M., Gennari, P., and Gualtieri, V. (2016) Measuring the results of FAO's statistics work. Proceedings ICAS VII Seventh International Conference on Agricultural Statistics I Rome 24-26 October.

Dupriez, O. (2018), An empirical comparison of machine learning classification algorithms, World Bank, http://pubdocs.worldbank.org/en/666731519844418182/PRT-OD-presentation-V2.pdf

Grimmer J., and Stewart, BM. (2013), "Text as data: The promise and pitfalls of automatic content analysis methods for political texts", *Political Analysis* Vol. 21, No. 3 (Summer 2013), pp. 267-297.

Hansen, S., McMahon, M., and Prat, A. (2018), Transparency and Deliberation within the FOMC: A Computational Linguistics Approach, Quarterly Journal of Economics, August.

Klein, T., Galdin, A., and Mohamedou, EL. (2016), "An indicator for statistical literacy based on national newspaper archives", in *Proceedings of the Roundtable Conference of the International Association of Statistics Education* (IASE), July 2016, Berlin, Germany.

Lowe, W. (2008), Understanding wordscores. *Political Analysis*, 16(4), 356-371.

McKenzie, D., and Sansone, D. (2017), Man vs. Machine in Predicting Successful Entrepreneurs Evidence from a Business Plan Competition in Nigeria.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., and Giovannini, E. (2005), *Handbook on Constructing Composite Indicators: Methodology and User Guid*e, OECD

Nardo, M., Saisana, M., Saltelli, A., and Tarantola, S. (2005), *Tools for Composite Indicators Building,* European Commission

PARIS21 (2015), "A scoring system to measure the use of statistics in the policy-making process", PARIS21, Paris, www.paris21.org/sites/default/files/Scoring_System_Use_Of_Data_2015_DFID.doc

Parthasarathy, R., Rao, V., and Palaniswamy, N. (2017), Deliberative Inequality A Text-As-Data Study of Tamil Nadu's Village Assemblies. World Bank Policy Research Working Paper No. 8119. Washington D.C.

Russell, M., and Muñoz-Ayala, J. (2015), Un estudio exploratorio para medir el uso de estadísticas en el diseño de política pública (Working Paper No. IDB-BP-374) Washington, D.C.: BID.

Scott, C. (2005), "Measuring up to the measurement problem: the role of statistics in evidence-based policy-making", PARIS21, Paris, www.paris21.org/node/672

UNECE. (2017, June), Extract of the recommendations on promoting, measuring and communicating the value of official statistics. Note by the Task Force on Value of Official Statistics (Rep. No. ECE/CES/2017/4). https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2017/CES_4_E_Value_of_official_stats.pdf

UNSD. (2016), Integrating a Gender Perspective into Statistics. Studies in Methods, Series F No. 111 https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Handbooks/gender/Integrating-a-Gender-Perspective-into-Statistics-E.pdf

UN Women (2018), Transforming Promises into Action: Gender Equality in the 2030 Agenda for Sustainable Development https://www.unwomen.org/en/digital-library/publications/2018/2/gender-equality-in-the-2030-agenda-for-sustainable-development-2018

DATABASES

Freedom House (2017), *Freedom of the press*. Freedom House. Retrieved from: https://freedomhouse.org/report-types/freedom-press

OECD (2017), *OECD.Stat*, Organization for Economic Cooperation and Development. Retrieved from: https://stats.oecd.org/Index.aspx?ThemeTreeID=3&lang=en

Scartascini, C., Cruz. C., and Keefer, P. (2017), *Database of Political Institutions*, Inter-American Development Bank

UNDP (2016), *Human Development Reports*. United Nations Development Programme. Retrieved from: http://hdr.undp.org/en/data

UNESCO UIS (2017), Adult literacy rate, United Nations Education Scientific and Culture Organization. Retrieved from: http://data.uis.unesco.org/index.aspx?queryid=166

World Bank Group (2017), *Country Policy And Institutional Assessment*. Retrieved from: https://datacatalog.worldbank.org/dataset/country-policy-and-institutional-assessment

World Bank (2017), *Statistical Capacity Indicators,* The World Bank. Retrieved from: http://databank.worldbank.org/data/reports.aspx?source=Statistical-capacity-indicators

World Bank (2017), *World Development Indicators*, The World Bank. Retrieved from: https://data.worldbank.org/products/wdi

World Bank (2017), *World Governance Indicators*, The World Bank. Retrieved from: https://datacatalog.worldbank.org/dataset/worldwide-governance-indicators

ANNEXES

ANNEX 1: RESULTS FROM MEASURING THE REFERENCES TO GENDER STATISTICS IN POLICY DOCUMENTS

Document sets

Around 134 policy documents, mostly national development plans, poverty reduction strategies and gender policy documents were analysed to assess the use of gender statistics in policy documents. The analysis was developed using a gender-specific keyword list, which included language from documentation of all gender-related SDG indicators on the UN website.

Text mining, generating and updating the keyword list

PARIS21, in collaboration with UN Women, expanded the universal coverage to a dimension that focuses on the use of gender statistics in key policy documents. To reach this goal, we added another layer of the filter after the detection of statistical terms. The new filter searches for words related to gender in the sentence or paragraphs detected in the previous identification of statistical terms. The new methodology also added another search engine to search for the 53 SDG indicators that have an explicit gender dimension.[6]

There are two sets of keywords used in this process. The first set contains a list of disaggregation keywords, and the second contains the gender-related SDG keywords. The methodology and keyword lists developed by PARIS21 and UN Women were validated in October 2019, following an Expert Meeting.[7] For the list of gender-specific vocabulary, please consult Annex 3 (section "Gender") of this document.

There are other methodology improvements made to develop the gender dimension. The stemming and stop words removal (see Section 2.a in the main text) were reviewed carefully by the authors to avoid loss of key information. The authors built the final stop words list based on the *Snowball* stop words list,[8] one of the most frequently used stop words lists developed in Porter (2001), with several modifications. For example, the word "ownership", which becomes "own" after stemming and later removal as a stop word, should be kept because of its significant correlation with gender statistics on women's economic status (linked to asset ownership).

The data source has several limitations that are successfully addressed. First and foremost, our hierarchy of statistical thinking in three stages of use creates a scale that has widely been validated, empirically, as a measure of statistical literacy. Nevertheless, the indicator measures a count of terms specifically referring to each level of use, whereas use would also need to be tested against the "appropriateness" of the terms used, in context. Therefore, the measure is conditional on the assumption that statistical terms are appropriate for the context in which they are used. This assumption is essential to a fully automated process allowing for the daily collection and analysis of policy documents.

The methodology also aims to ensure that the statistical keywords used in the data mining methodology correspond to the public documents in the sample. For this, and following the nascent literature, a new method is implemented to define the list of statistical terms that will feed the

[6] *The selection of these indicators is based on UN Women (2018) and takes into account the conclusions of the IAEG-SDG https://unstats.un.org/sdgs/iaeg-sdgs/report-iaeg-sdgs/*

[7] *https://paris21.org/news-center/news/PARIS21-UN-Women-Gender-Statistics-Expert-Meeting*

[8] *The list can be found in The stop word lists can be found in http://snowball.tartarus.org/dist/snowball_all.tgz.*

indicator (Grimmer and Stewart, 2013) based on the generic indicator (outlined above in the main analysis). The algorithm, known as "term-frequency–inverse document frequency" (tf-idf), is widely used in text mining practices. This was used for identifying sectoral approaches that are present in general policy documents, as well as capturing how monitoring and evaluation arrangements are introduced and how the results of previous plans are discussed. Policy documents and SDG documents related to gender were used for the machine to "learn" the pattern of sectoral policy text in policy documents.

Document sets

The text mining analysis was applied to 134 PRSPs and NDPs belonging to 77 countries that were written in English, French and Spanish and available on the internet.

Table A1 Distribution of countries and the number of documents across continents

Continent	Countries	Documents
Asia	20	40
Africa	35	65
Americas	10	13
Europe	4	5
Oceania	8	11

The period covered by the indicator spans the years 2001 to 2020. Because countries follow different timetables for sanctioning these plans, two waves were devised: the first between 2001 and 2010 and the second between 2011 and 2020. The tables below (A.4.2 and A.4.3) provide more details on the number of documents per wave. There are policy documents for 55 countries distributed in two waves in the corpus of documents. Documents available for 43 countries in both waves are included for panel comparison purpose.

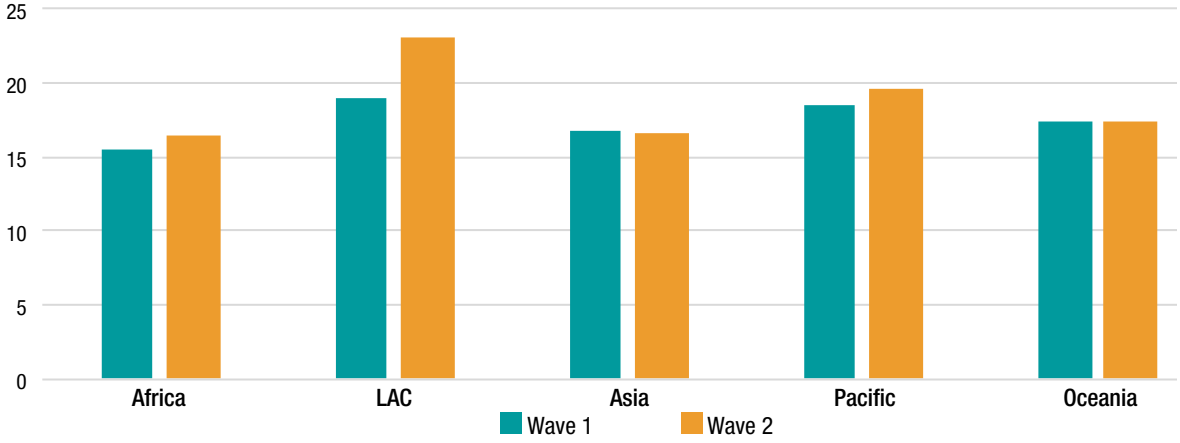
Table A2 Number of documents in each wave

	2001 – 2010	2011 – 2020
Number of countries with 1 document	40	71
Number of countries with 2 documents	15	10

Preliminary results

The preliminary results show that in 55 countries with at least one policy document in both decades of this century, the use of gender statistics increased slightly in the last decade (2011-2020). The Americas and Europe lead the way in the use of gender statistics, while Asian countries have a higher average score in quoting actual numbers. Figures below present the UoS scores for the second wave and the difference between the average scores for waves 1 and 2, by region.

Figure A1



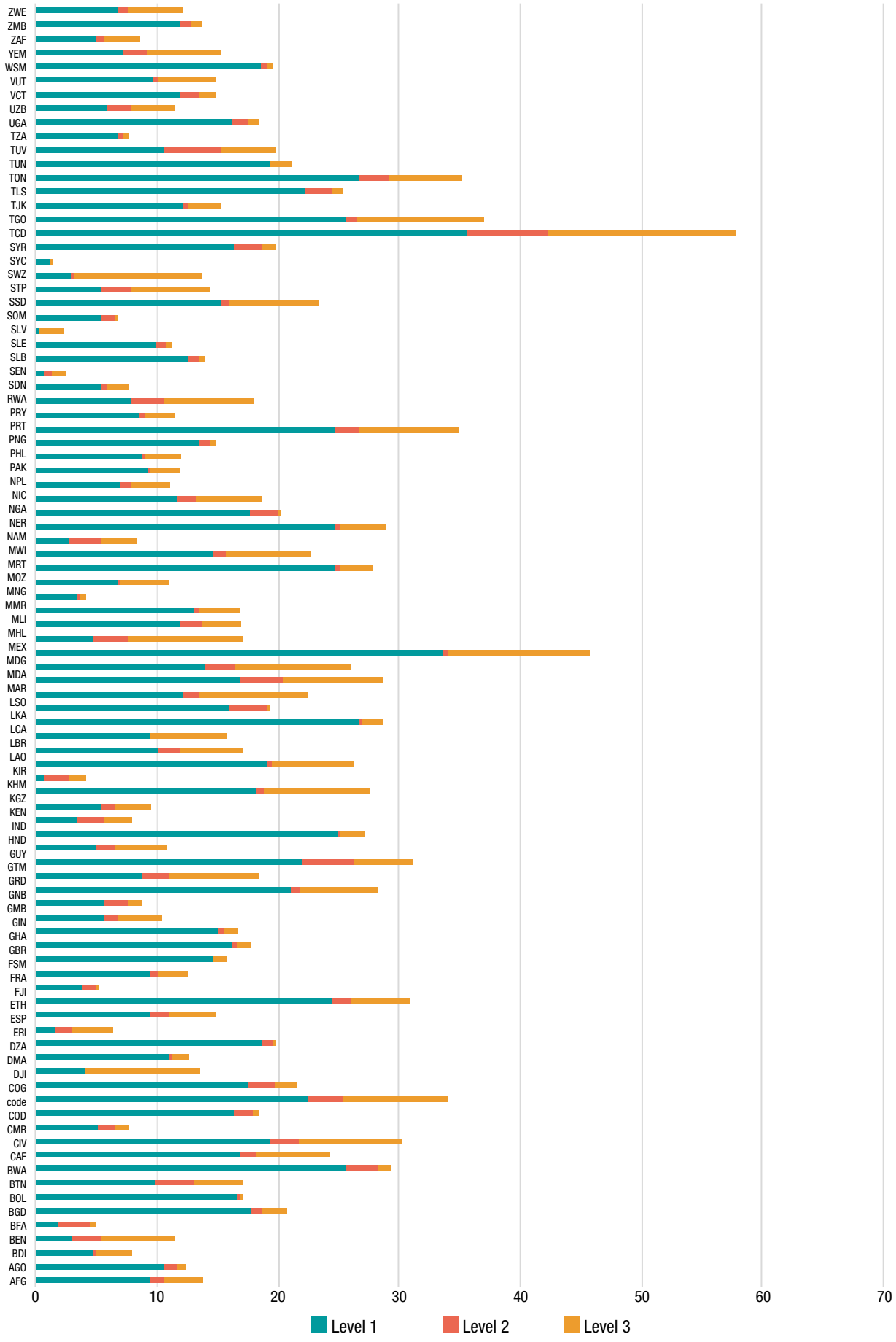
The positive correlation between country scores in the two waves was relatively low. The lack of a strong correlation suggests that the progress on the use of statistical evidence for policymaking was not linear, which is consistent with the finding for general use of statistical terms. While this volatility can be partly explained by the shifts in policy priorities linked to political dynamics, there is also a possibility that the adoption of the Agenda 2030, as well as the increased use of data following the data revolution, also contributed to this result. Further analysis is required, but these results likely indicate a need to continue to develop the supply of gender statistics as well as for continued efforts to advocate for gender-sensitive approaches to development policy design, in order to advance gender equality.

Limitations

The adapted gender-specific methodology inherited several limitations of the generic methodology. One specific limitation is that the indicator measures a count of terms referring to each level, which would also need to be tested against the “appropriateness” of the context in which they were used. The validity of the scores is conditional on the assumption that the keywords are suitable for the context in which they are situated. This assumption is essential to a fully automated process for analysing policy documents.

The current sample size of the exercise (i.e., number of available policy documents in the two waves) limited the possibility of finding more information using the current methodology. As the overall frequency of using gender statistics is much lower

than using generic statistical terms, the inclusion of more documents in the sample would help smooth the volatility of the score. Moreover, with an increased number of countries starting to develop gender-specific policy documents, it will be helpful to include those documents for further analysis of how statistical evidence is used in gender-responsive policymaking, in order to address gender equality.


ZWE
ZMB
ZAF
YEM
WSM
VUT
VCT
UZB
UGA
TZA
TUV
TUN
TON
TLS
TJK
TGO
TCD
SYR
SYC
SWZ
STP
SSD
SOM
SLV
SLE
SLB
SEN
SDN
RWA
PRY
PRT
PNG
PHL
PAK
NPL
NIC
NGA
NER
NAM
MWI
MRT
MOZ
MNG
MMR
MLI
MHL
MEX
MDG
MDA
MAR
LSO
LKA
LCA
LBR
LAO
KIR
KHM
KGZ
KEN
IND
HND
GUY
GTM
GRD
GNB
GMB
GIN
GHA
GBR
FSM
FRA
FJI
ETH
ESP
ERI
DZA
DMA
DJI
COG
code
COD
CMR
CIV
CAF
BWA
BTN
BOL
BGD
BFA
BEN
BDI
AGO
AFG
0
10
20
30
40
50
60
70
Level 1
Level 2
Level 3

ANNEX 2: LIST OF INCLUDED DOCUMENTS (PRSPS AND NDPS)

Country	Year
Afghanistan	2008
Afghanistan	2017
Albania	2001
Albania	2013
Armenia	2001
Armenia	2003
Armenia	2008
Armenia	2014
Azerbaijan	2003
Azerbaijan	2008
Azerbaijan	2017
Bahamas	2017
Bangladesh	2005
Bangladesh	2009
Bangladesh	2011
Bangladesh	2016
Barbados	2006
Barbados	2013
Belize	2010
Benin	2003
Benin	2007
Benin	2011
Bhutan	2002
Bhutan	2004
Bhutan	2008
Bhutan	2009
Bhutan	2013
Bolivia	2001
Bolivia	2016
Bosnia and Herzegovina	2004
Botswana	2003
Botswana	2009
Burkina Faso	2004
Burkina Faso	2011
Burundi	2006
Burundi	2012
Cambodia	2001
Cambodia	2005
Cambodia	2009
Cambodia	2014
Cameroon	2003
Cape Verde	2004
Cape Verde	2008
Central African Republic	2007
Central African Republic	2017
Chad	2008
Chad	2017
Lesotho	2005
Lesotho	2012

Country	Year
Comoros	2005
Comoros	2009
Côte d'Ivoire	2009
Côte d'Ivoire	2012
Democratic Republic of the Congo	2006
Democratic Republic of the Congo	2013
Djibouti	2004
Djibouti	2015
Dominica	2006
Egypt	2016
Eritrea	2004
Ethiopia	2002
Ethiopia	2006
Gambia	1994
Gambia	2006
Gambia	2012
Gambia	2017
Georgia	2003
Georgia	2009
Ghana	2005
Ghana	2010
Grenada	2006
Guinea	2002
Guinea	2007
Guinea	2011
Guinea	2013
Guinea-Bissau	2006
Guinea-Bissau	2011
Guyana	2002
Haiti	2007
Haiti	2014
Honduras	2001
India	2008
Indonesia	2003
Indonesia	2010
Iraq	2010
Iraq	2013
Jordan	2013
Kenya	2004
Kenya	2008
Kuwait	2015
Kyrgyzstan	2007
Kyrgyzstan	2012
Lao PDR	2004
Lao PDR	2006
Lao PDR	2011
Lao PDR	2016
Liberia	2008
Madagascar	2007

Country	Year
Malawi	2002
Malawi	2006
Malawi	2012
Malaysia	2016
Maldives	2007
Mali	2006
Mali	2013
Mauritania	2000
Mauritania	2006
Mauritania	2011
Micronesia	2004
Moldova	2004
Moldova	2008
Moldova	2012
Mongolia	2003
Mongolia	2007
Mongolia	2016
Mozambique	2004
Mozambique	2006
Namibia	2002
Namibia	2007
Namibia	2012
Namibia	2017
Nepal	2002
Nepal	2003
Nepal	2007
Nicaragua	2009
Niger	2002
Niger	2007
Niger	2012
Nigeria	2004
Nigeria	2017
Niue	2009
Oman	2016
Pakistan	2001
Pakistan	2003
Pakistan	2009
Palestine	2014
Papua New Guinea	2004
Papua New Guinea	2010
Philippines	2004
Philippines	2017
Qatar	2008
Qatar	2011
Republic of Congo	2008
Republic of Congo	2012
Rwanda	2007
Rwanda	2013
Samoa	2008
Samoa	2012
Sao Tome and Principe	2003
Sao Tome and Principe	2005
Saudi Arabia	2005
Saudi Arabia	2010
Senegal	2002
Senegal	2006
Senegal	2013
Serbia	2004
Seychelles	2000
Seychelles	2012
Sierra Leone	2005
Sierra Leone	2008
Sierra Leone	2013
Solomon Islands	2016
Somalia	2017
South Korea	2009
South Sudan	2011
Sri Lanka	2017
Sudan	2007
Sudan	2012
Swaziland	2005
Swaziland	2006
Swaziland	2013
Tajikistan	2002
Tajikistan	2010
Tajikistan	2013
Tanzania	2005
Tanzania	2010
Tanzania	2011
Tanzania	2016
Timor-Leste	2002
Timor-Leste	2011
Togo	2009
Tonga	2010
Trinidad and Tobago	2016
Turkey	2001
Turkey	2007
Turkey	2014
Tuvalu	2005
Uganda	2015
United Arab Emirates	2011
Uruguay	2016
Uzbekistan	2007
Vanuatu	2006
Vietnam	2003
Yemen	2002
Zambia	2006
Zambia	2011

ANNEX 3: LIST OF KEYWORDS

Health

Abortion
Affordable medicines
AIDS incidence
AIDS prevalence
Anemia
Antenatal care
Antiretroviral therapy coverage
Birth rate
Birth registration
Births attended by skilled personnel
Birthweight
BMI
Body mass index
Cancer
Cardiovascular diseases
Cause of death
Chronic respiratory disease
Communicable diseases
Community health workers
Contraceptive prevalence
Death registration
Density of medical personnel
Diabetes
Diarrhoea
Drug abuse
Food poisoning
Harmful substance abuse
Health insurance coverage
Health worker density
Health worker distribution
Hepatitis B
HIV incidence
HIV prevalence
Hospital beds
Immunisation
Improved sanitation facilities
Insecticide treated nets
Insufficient physical activity
ITN
Life expectancy
Malaria
Malnutrition
Maternal mortality
Mental health disorders
Mother-to-child transmission
Neglected tropical disease
Neonatal mortality
New HIV infections
Non-communicable diseases
Nurses
Out-of-pocket payment
Overweight
Physicians
Pneumonia
Population growth
Postpartum care
Raised blood pressure
Reproductive health
Salt intake
Sero-positivity
Sex ratio
Sexually transmitted infections
Specialist surgical workforce
STI
Stillbirth rate
Stunting
Suicide rate
TB
Traffic injuries
Treatment interventions
Tuberculosis
Under-five mortality
Undernourishment
Underweight
Unsafe sanitation
Vaccine preventable diseases
Wasted

Education and skills

Adolescents out of school
Adult illiterates
Children not in school
Children out of school
Completion rate
Drop out
Dropout
Educational attainment
Emigration rate
Enrolment
Government expenditure on education
Graduation ratio
Gross enrolment ratio
Gross intake ratio
Household Expenditure on education
ICT skill
Illiteracy
Literacy rate
Minimum proficiency in mathematics
Minimum proficiency in reading
Net enrolment rate
Net intake rate
Numeracy
OOS
Over-age students
Persistence to grade
Private enrolment
Promotion rate
Pupil-teacher ratio

Agriculture and Livestock

Agricultural export subsidies
Agricultural irrigated land
Agricultural land
Agricultural machinery
Agriculture orientation index
Agriculture production
Annual freshwater withdrawals
Arable land
Area under sustainable agriculture
Cereal production
Cost to export
Cost to import
Crop production
Documents to export
Documents to import
Employment in agriculture
Export value
Export volume
Fertiliser consumption
Food exports
Food imports
Food loss
Food price anomalies
Food production

Import value
Import volume
Income of small-scale producers
Irrigated land
Land area
Land under cereal production

Lead time to export
Lead time to import
Livestock production
Livestock production
Permanent cropland
Producer Support Estimate
Production per labour unit
Raw materials exports
Raw materials imports

Industry

Arms exports
Burden of customs
Business density
Capital efficiency
Changes in inventories
Competition
Construction
Cost to export
Cost to import
Delay in obtaining electrical connection
Documents to export
Documents to import
Ease of doing business

Employers
Employment in industry

Enterprise density
Export value
Export volume

Firms
Firms offering training
Firms using banks
Fixed capital formation
Formally registered firms
Global Value Chains
Goods expense
Goods exports
Goods imports
Gross capital formation
Gross output
ICT goods exports
ICT goods imports
Import value
Import volume

Industry sector contribution
Innovation

Labor contributions
Labor tax
Lead time to export
Lead time to import
Losses due to theft, robbery, vandalism, and arson
Manufactured exports
Manufactured imports
Manufacturing sector contribution
Manufacturing value added
Medium and high-tech exports
Private credit
Private sector investment

Procedures to build a warehouse
Procedures to register property
Procedures to start a business
Process
Productivity
Profit tax
Public credit
Recruitment cost
Research and Development
Sectors
Small-scale industries
Smes

Structural Change
Structural Transformation
Tariff rate

Tax burden
Tax payments

Tax rate
Technology
Time dealing with requirements of government regulations
Time to build a warehouse
Time to clear exports
Time to enforce a contract
Time to get electricity
Time to obtain a license
Time to prepare and pay taxes
Time to register property
Time to start a business
Trade in goods
Unregistered firms
Upgrading

Services

Access to credit
Access to finance

Access to financial services
Binding coverage

Cash transfer
Commercial bank branches
Commercial banks
Cost of business start-up
Employment in services

Firms using banks
Formally registered firms
ICT service exports
ICT service imports
Insurance and financial services
Exports
Insurance and financial services imports
Labor tax

Access to credit
Access to finance
Access to financial services
Binding coverage
Cash transfer
Commercial bank branches
Commercial banks
Cost of business start-up
Employment in services
Firms using banks
Formally registered firms

ICT service exports
ICT service imports
Insurance and financial services
Exports
Insurance and financial services imports

Labor tax
Micro finance
Microcredit
Mobile account
Access to credit
Access to finance
Access to financial services
Binding coverage
Cash transfer
Commercial bank branches
Commercial banks
Cost of business start-up
Employment in services
Firms using banks
Formally registered firms
ICT service exports
ICT service imports
Insurance and financial services
Exports
Insurance and financial services imports
Labor tax
Micro finance
Microcredit
Mobile account
Access to credit
Access to finance
Access to financial services
Binding coverage
Cash transfer
Commercial bank branches
Commercial banks
Cost of business start-up
Employment in services
Procedures to enforce a contract
Procedures to get electricity
Procedures to obtain a license
Gasoline price
GDP per unit of energy
Generation capacity
Metals exports
Metals imports
Mineral rents
Mineral rents
Mining sector contribution
Natural gas rents
Natural gas rents
Oil rents
Oil reserves
Ores exports
Ores imports
Quarry
Renewable energy consumption
Value lost due to electrical outages

Gender

Adolescent fertility rate
Contraceptive
Domestic violence
Family care work
Family planning
Female construction graduates
Female employer
Female engineering graduates
Female genital cutting
Female genital mutilation
Female judges
Female manufacturing graduates
Female ownership
Female police officers
Female science graduates
Females among tertiary education teachers
Fertility rate
FGM
Gender gap
Gender parity index
GPI
Intimate partner violence
Lifetime risk of maternal death
Physical violence
Prenatal care
Seats held by women
Seats held by women in parliament
Sexual violence
Unpaid domestic work
Micro finance
Microcredit
Mobile account
Women marred before age 18
Women with secure tenure of land

Environment

Access to clean fuels
Air pollution
Ammonia
Bird species
CO2 emissions
CO2 intensity
Deforestation
Degradation
Degraded land
Disaster economic loss
Disaster resilient habitats
Disaster risk reduction
Disaster risk reduction
Domestic material consumption
Droughts
Droughts extreme temperatures
Electricity production
Energy use
Eutrophication
Extreme temperatures
Fine particulate matter air pollution
Fish farming
Fish species
Fish stocks
Fisheries production
Floating plastic debris
Floods
Forest area
Forest coverage
GHG net emissions
Greenhouse gas emissions
Hazardous waste
HFC gas emissions
HFC gas emissions
Illegally poached wildlife
Illicitly trafficked wildlife
Land area
Land consumption
Mammal species
Marine acidity
Marine protected areas
Material footprint

Material recycled
Methane emissions
Firms using banks
Formally registered firms
ICT service exports
Nitrogen oxides
Nitrous oxide emissions
NO2
NH3
Nitrogen oxides
Nitrous oxide emissions
NO2
NOX
Plant species
PM2.5
Population growth
Precipitation
Recycling rate
Red List Index
Reforestation
Renewable electricity output
Sales of pesticides
Shelter
SO2
Sulphur oxides
Terrestrial protected areas
Sales of pesticides
Shelter
Wastewater treatment
Water quality
Water stress
Water-use efficiency
WUE

Employment and social inclusion

Absolute poverty
Allowances
Benefits incidence in poorest quintile
Body weight
Caloric intake

Compensation of employees
Contributing family workers
Coverage of social insurance programs
Coverage of social safety net programs

Coverage of unemployment benefits
Crowding
Dependency ratio
Depth of poverty
Disability
Disability benefit
Disposable income
Earnings inequality
Employment status

Women in employment in the non-agricultural sector
Women in managerial positions
Expenditure on housing
Female employment in senior and middle management
Female to male labor force participation
Food insecurity
Food reserve
Food security
Formal employment

GDP per person employed
Generalised entropy ratio
GINI
Growth in poverty
Hourly earnings

Household Consumption

Household final consumption expenditure
Household Income
Human Development Index
Incidence of poverty
Informal employment
International poverty line
Labor force participation rate
Legal title to agricultural land
Minimum level of dietary energy
National poverty line
NEET
Net migration
Occupational injury frequency
Own account
Part-time labor
Perceived poverty
Persons per room
Poverty gap

Poverty headcount
Precarious employment
Private consumption

Proportion of poor households
Race
Real wages
Reduction in poverty
Relative poverty
Rural poverty
Salaried workers
Self-employed
Severely poor households
Severity of poverty
Social empowerment
Social mobility
Social security

Urban poverty
Vulnerable employment

Wage gap

Work participation
Working hours of children
Working poor
Youth not in education, employment or training

Housing and public service delivery

Access to public transport
Access to sanitation
Active mobile-broadband subscriptions
Area where elevation is below 5 meters
Armed forces personnel

Drinking water
Fixed broadband subscriptions
Fixed Internet broadband
Fixed-telephone subscriptions
Homeless persons
Homelessness
Households with a computer
Households with Internet access
Improved water source
Individuals using the Internet
Investment in energy
Investment in telecom
Investment in transport

Investment in water and sanitation
Mobile cellular subscriptions
Mortality by road traffic injury
Population density
Population in largest city
Population in slums
Population in urban agglomerations of more than 1 million
Population living below 5 meters
Power outages
Safe water
Sanitary facilities
Satisfaction with public service
Unsafe water
Urban population
Waste collection
Waste treatment
Water supply

Infrastructure and transportation

Access to services
Employment to population ratio
Expenditure on food
Bridges
Electric power consumption
Freight volumes
Gravel road
Internet servers
Maintenance
Paved roads
Police
Postal service
PPP
Procurement
Maintenance
Public private partnerships
Quality of port infrastructure
Radio
Rail lines
Railways
Road network
Shipping
Television
Traffic
Waterways

Culture and sports

Archive
Arts
Craft
Culture
Dancing
Folk
Heritage
History
Language
Library
Literature
Music
Painting
Religion
Tribe
Archive
Arts
Craft
Culture
Dancing
Folk
Heritage
History
Subsistence worker
Underemployment
Unemployment rate
Music
Painting
Religion
Tribe

Public finance and macroeconomic stability

Acquisition of financial assets
Adjusted net national income
Adjusted net savings
Adults with an account
Agriculture value added
Average grace period
Average maturity
Balance of payments
Bank capital to assets ratio
Bank liquid reserves
Bonds
Borrowing
Borrowing
Broad money
Budget deficit
Capital account
Capital-labour ratio
Central government debt
Claims on central government
Commitments
Concessional debt
Consumer price index
CPI
Credit-to-GDP ratio
Currency composition
Current account balance
Current transfers
Customs and other import duties
Debt forgiveness
Debt service
Debt stock reduction
Debt-to-export ratio
Debt-to-GDP ratio
Deposit interest rate
Disbursements
Domestic budget
Domestic credit
Expense
Export market penetration
Export performance
Air transport
Airport
All-season road
External debt stocks
Financial account
Fiscal deficit
Foreign direct investment
GDP
GDP deflator
GNI
Goods and services expense
Government final consumption expenditure
Grants
Gross Domestic Product
Gross domestic savings

Gross fixed capital formation
Gross national expenditure
Gross National Income
Gross savings
Gross value added
Income share
Incurrence of liabilities
Incurrence of liabilities

Inflation
Interest arrears
Interest forgiven

Interest payments
Investment in nonfinancial assets
Lending
Lending interest rate
Market capitalisation
Multilateral debt
Net bilateral aid flows

Net financial flows
Net flows on external debt
Net official development assistance (ODA)
Net official flows from UN agencies
Net trade
Net transfers on external debt
Nonperforming loans
Personal remittances
Personal remittances
Portfolio equity
Portfolio Investment
Price level ratio
Primary income

Real effective exchange rate
Real interest rate
Remittance costs

Language
Library
Literature
Social contributions
Spending on education
Spending on essential services
Spending on health
Spending on social protection
Subsidies

Tax compliance
Tax revenue
Taxes on exports

Taxes on goods and services
Taxes on income, profits and capital gains
Taxes on international trade
Taxes on products
Technical cooperation
Terms of trade
Total reserves
Trade
Wholesale price index

Science, Technology and Innovation

Patent applications
R&D expenditure
Research and development expenditure
Researchers in R&D
Scientific journal articles
Technicians in R&D
Trademark applications

Governance

Access to government information
Access to public information
Accountability

Affected by disaster
Arbitrary detention
Arms exports
Battle-related deaths
Bribery
Children in employment
Civil liberties
Conflict-related deaths
Corruption
Detained prisoners without sentence
Discrimination
Forced evictions
Forced labour

Exports

External balance
Human rights
Human trafficking

Illicit financial flows
Intentional homicides
Internal refugees
Internally displaced persons
Law enforcement officials killed in duty
Length of pre-trial detention
Military expenditure
Missing persons
Physical and non-physical abuse
Physical punishment
Political rights
Population feeling unsafe
Population with secure tenure
Psychological aggression
Refugee population
Registered births
Rule of law
Statistical Capacity
Transparency
Unlawful detentions
Unsentenced detainees
Violent crimes
Voter turnout
Human rights
Human trafficking

Illicit financial flows
Intentional homicides
Internal refugees
Internally displaced persons
Law enforcement officials killed in duty
Length of pre-trial detention
Military expenditure
Missing persons
Physical and non-physical abuse
Physical punishment
Political rights
Population feeling unsafe
Population with secure tenure
Psychological aggression
Refugee population
Registered births
Rule of law
Statistical Capacity
Transparency
Unlawful detentions
Unsentenced detainees

Violent crimes
Voter turnout

Resources allocated to poverty reduction
Revenue

Statistical Terms

Model
Utility
Outcome
Multinomial
Logit
RPL model
Estimator
Aggregate
Cohort-level
Nonlinear
Estimator
Covariance
Variance
Residual
Sampling
IV
Instrumental
Kernel
Probit
Poisson
Panel
Quantile
Survival
Treatment
Arellano-bond
Bias

Spurious
Error
Observations
Relationship
Confounding
Non-random
Randomisation
Effect
Hypotheses
Significance
Power
Omitted
Autoregressive
Asymptotic
Fixed Effects

First-differences
Central Limit
Stochastic
Autocorrelation
Upward
Downward
Matching
Difference-in-Difference
Bayes

Hazardous child labour

Hearings
Endogenous
Misspecification
Imputation
Semiparametric
Mean
Mode
Median
Probability

Aggregate
Marginal
Confidence Interval
P-value
Heteroskedastic
Time series
Parameter
Rescaling
Granger
Census
Multivariate
Cluster
Chi-square
Dummy
Deviation
Coefficient
Likelihood
Conditional
Ceteris-paribus
Cumulative
Regression
Data set
Data mining
Cross-section
Heckman
Heterogeneity

Residual
Hot deck

Cold deck
Robust
Unbalanced
Longitudinal
Counterfactual
Propensity
Exogenous
Measurement
Microdata
Regressor
Multicollinearity
Attrition
Odds ratio
Polynomial
Propensity score
Quasi-experiment

Treatment group
Control group
Unobserved
Average
BLUE
OLS

Disaggregation

region
country
district
locality
neighbourhood
province
suburb
territory
zone
south
east
west
north
department
community
rural
agrarian
agricultural
farm
land
urban
city
town
village

gender
male
female
girl
boy
men
women
man
woman
sex
age
old
young
elderly
youth
adolescence
adolescent
adult
child
Linear
Tobit
Best Linear Unbiased Predictor

Reference to previous plan:

The previous strategy
The previous plan
Previous PRSP
Previous NDP
The previous Vision
Implementation report
Reviewing performance
Evaluation of the Implementation
Lessons learnt
Development during
Assessment of implementation
Review of implementation
The last years
The first years
Previous years
The plan aimed
Development yield
Situation analysis
Achievements + past tense
Challenges + past tense
Treatment effect
Inference
Treatment evaluation

Monitoring and Evaluation

Monitoring
Evaluation
Performance
Framework
Results-based
Institutional arrangement
Progress
Follow-up
M&E

ANNEX 4: EXAMPLES OF SENTENCES IDENTIFIED FOR EACH COMPONENT FROM PHILIPPINES' 2017 NATIONAL DEVELOPMENT PLAN

Basic Use

(green indicates level 1 use; red indicates level 2 use; purple indicates level 3 use).

Level 1:

Likewise, population with little to no education also have poorer health outcomes because they lag behind in nutrition and health indicators.

Level 2:

Growth in global trade (export volume) also slowed significantly from an average of 8.3 percent in 2003-2007 to 3.0 percent in 2008-2015.

Drop-out rates decreased from 6.29 percent in SY 2010-2011 to 2.70 percent in SY 2015-2016 at the elementary level and from 7.79 percent to 6.65 percent at the secondary level.

Level 3:

Baseline figures for goods and services exports are based on BOP (BPM6) data.

Figures for 2017 – 2022 were estimated based on annualised 2016 BPM6 levels and latest DBCC assumptions on growth rates for exports of goods and services approved on 20 December 2016

Disaggregation

Labor force participation of women has improved only slightly over the past 25 years, from about 47 percent in the early 1990s to around 50 percent in more recent years

The 2012 basic sector data from the Philippine Statistics Authority (PSA) noted that 35.2 percent of Filipino children are poor.

In 2011, the number of working children was estimated at 3.3 million, of which 2.1 million were engaged in child labor.

Monitoring and evaluation arrangements

The government will formulate the framework and its subsequent indicators, consistent with the Philippine Statistical Development Plan (PSDP) and aligned with the UN Social Development Goals (SDGs) to deliver quality SP statistics in support of evidence-based policymaking, program implementation, and monitoring and evaluation.

For monitoring and evaluation, conduct of periodic nationwide Victimisation Survey covering crimes and human rights violations will be pursued to augment administrative-based data.

Assessment of previous plans

Sustaining growth at the rate of 6.3 percent during the previous plan period, the economy is now clearly on a higher growth trajectory.

However, a backlog of 66,463 classrooms remains from the previous plan period largely due to transferees from private schools.


DATA
DATA
Progress
Growth & Decline


JOHN
SAM
LISA
10.8%
7.0%
GLOBAL SHARE

Available at: www.paris21.org/xxxxx
Contact the PARIS21 Secretariat at contact@paris21.org
Visit www.PARIS21.org
Follow PARIS21
@contactparis21
/company/paris21/