Received: 15 October 2020 | Revised: 27 May 2021 | Accepted: 16 June 2021

DOI: 10.1002/sd.2223

RESEARCH ARTICLE

Sustainable Development | WILEY

Monitoring sustainable development goals and the quest for high-quality indicators: Learning from a practical evaluation of data on corruption

Giulia Mugellini | Jean-Patrick Villeneuve | Marlen Heide

Institute of Communication and Public Policy (ICPP), Università della Svizzera Italiana (USI), Lugano, Switzerland

Correspondence
Giulia Mugellini, Institute of Communication and Public Policy (ICPP), Università della Svizzera Italiana (USI), via G. Buffi 13, 6900 Lugano (Switzerland).
Email: giulia.mugellini@usi.ch

Abstract

Tracing progress in implementing the sustainable development goals (SDGs) is at the core of pushing and accounting for change. However, monitoring SDGs is challenged by a lack of purpose-fit and high-quality indicators based on data that are collected through a sound methodology, generated regularly, comparable over time, and publicly accessible. Assessing and improving the quality of existing data is essential for helping countries to generate an evidence base for action. General criteria for evaluating data quality are already available at the national and international level but their practical operationalization for the assessment of specific SDGs indicators is still underdeveloped. Taking target 16.5 as a case study, this paper evaluates the quality of existing corruption surveys and their relevance for SDGs. Results show that the main challenges concern data validity (they measure only one aspect of corruption), comparability (they use culturally biased definitions), periodicity (they are not regularly developed), and raw-data accessibility. This paper develops an original framework for benchmarking the overall methodological quality of existing corruption metrics. This framework can be used beyond the immediate context of corruption measurement and SDGs assessment. The same logic and methodology can, indeed, be employed to evaluate the quality of other metrics and support national governments and practitioners in identifying the informational and methodological gaps to be addressed in order to improve and make the best use of available statistical information.

KEYWORDS
corruption, data availability, data quality, policy design, sample surveys, sustainable development goals

1 | INTRODUCTION

The United Nations Agenda for Sustainable Development[1] identifies 17 goals to be achieved by member states by 2030. In order to monitor the progress of the sustainable development goals (SDGs), the United Nations Statistical (UNStats) Commission approved a global framework of indicators (UN SDGs indicators) (Endnote 1).

The UN SDGs indicators are assessed by the international community according to their methodological development (whether the indicators have an established international methodology and standards), and data availability (to what extent data on SDGs indicators are available across Member States). The methodological development is periodically assessed by the UN itself (Endnote 1). Moreover, experts have taken stock of the data readily available for

SDGs assessment (see for example Allen et al., 2018; Eurostat, 2020; Sachs et al., 2020).

In addition, available data have been used to evaluate countries' performance with regard to the progress in SDGs implementation (Miola & Schiltz, 2019; Lafortune et al., 2018; Sachs et al., 2018; OECD, 2017; Eurostat, 2018).

However, the quality of available data used to monitor specific SDGs seems understudied. For countries to implement SDGs, it is critical to have an evidence-base for action (Allen et al., 2018) and to know the quality of this evidence. Yet, ensuring the quality of existing data enables countries to assess their performance, comply with inter-organizational cooperation requirements (Batini & Scannapieco, 2006), and develop more targeted policies (Chengalur-Smith et al., 1999)–within the context of the SDGs and beyond. Furthermore, the possibility of using sound existing data is normally more cost-effective than the collection of new ones.

While general criteria for evaluating data benchmarking SDG indicators are included in the United Nations National Quality Assurance Frameworks Manual for Official Statistics (United Nations, 2019), more practical applications of these generic quality dimensions to specific SDGs targets are still underdeveloped. Quality assessment frameworks addressing specific SDGs targets would be useful for national governments to identify informational and methodological gaps that need to be addressed to develop solid and sustainable monitoring systems.

Ensuring data quality is particularly challenging when measuring human behaviors and social phenomena. Their complexity and inferential nature make them difficult to be translated into functioning and measurable indicators and might lead to validity and reliability issues (Drost, 2011). SDG Target 16.5 represents an interesting case study for assessing the quality of data for SDGs, notably since it refers to a social phenomenon corruption–which is intrinsically difficult to define and measure.

The complexity of corruption in terms of definition and measurement has been widely recognized by scholars (Beeri & Navot, 2013; Bussell, 2015; Graycar, 2015; Gupta, 1995; Heywood, 2017; Heywood & Rose, 2014; Jancsics, 2019; Knack, 2007; Mungiu-Pippidi & Dadašov, 2016; Rose, 2018; Sequeira, 2012, UNODC & UNDP, 2018). Corruption is a multi-faceted phenomenon that involves different actors, behaviors and activities. The concept is broad and there is no consensus on a standardized, exhaustive definition. This creates a problem for the validity of corruption metrics (Rose, 2018; Thatcher, 2010), which requires that an indicator adequately reflects the concept to be measured.

Sample surveys are generally favored to measure experiences of corruption directly from those involved/impacted, be they citizens, businesses or civil servants, and thus overcome non-reporting problems affecting criminal justice statistics (Blind, 2011; Herrera et al., 2007; Jandl, 2017; Knack, 2007; Lynch, 2006; Sequeira, 2012; UNODC-UNDP, 2018). A major question is whether existing corruption surveys can be used to monitor Target 16.5. While some scholars have critically analyzed the method and content of international and European corruption surveys (Chabova, 2017; Malito, 2014; Wysmulek, 2019), national corruption surveys have so far not been evaluated neither with regard to their relevance for monitoring SDG Target 16.5, nor with regard to their quality.

Taking Target 16.5 as a case study, this paper evaluates the quality of existing surveys' data for measuring progress implementing the UN SDGs. Besides assessing baseline data for Target 16.5, this paper develops an original framework for benchmarking the overall methodological quality of existing corruption metrics considering validity, accuracy, comparability, periodicity, and raw-data accessibility. This framework can be also used by national government representatives and practitioners to assess and improve the quality of existing data collection on corruption. National institutional and international surveys have been analyzed and used to test and "calibrate" the quality assessment framework due to their methodologically accurateness and relevance for SDGs bench-marking. The testing of the framework and the analysis of existing corruption surveys might facilitate secondary data use by academics and practitioners. It might also support the ex-post harmonization of survey data necessary to compare surveys results across countries. Furthermore, it provides a repository of information on existing official corruption surveys and raises awareness among potential users of the existing data richness. On a more general level, the same logic and methodology used to develop this framework can be used to evaluate the quality of data and metrics referring to other phenomena, and thus support national governments and practitioners in making the best use of available statistical information.

2 | SCOPE, MATERIALS AND METHODS

The scope of this paper is to raise awareness on the importance of evaluating and improving the quality of existing data to generate an evidence base for action, and provide policymakers, practitioners and academics with the logic and methodology that can be used for assessing the quality of this evidence and its relevance for the UN SDGs. Data on corruption and UN SDGs Target 16.5 are taken as case study to achieve this scope.

In particular, this paper aims to:

1. Identify the international goals and methodological standards for data quality in general and for corruption measurements in particular, with a special focus on UN SDGs requirements (see section "Existing criteria for assessing data quality of SDGs indicators").
2. Develop an original framework that translates general criteria for data quality into more practical standards for assessing corruption measurements (see section "The Quality Evaluation Framework for SDGs target 16.5")
3. Test the applicability of the framework for corruption data quality on a set of existing surveys (see section "Testing the QUEST 16.5").

In order to fulfill the above-mentioned objectives, this paper follows a four-step methodology:

1. Literature review of the main international goals and methodological standards for the quality of data with a specific focus on UN SDGs requirements;

2. Literature review of international and methodological standards for the quality of corruption measurements;
3. Identification of the existing national and cross-national population and business corruption surveys developed by national statistical offices or other governmental agencies, by international agencies, and non-profit organizations;
4. Quality assessment of the content (e.g., types of corruption covered, formulation of questions and types of items included to gain further insights on corruption experience), and methodology (e.g., data collection method, sampling method) of the identified surveys.

The first two steps are necessary to build the framework, the third step is necessary to identify the set of surveys on which to test the framework, the fourth steps clarifies how the quality of the identified surveys has been assessed using the framework.

2.1 | Existing criteria for assessing data quality of SDGs indicators

All SDGs indicators are classified by the Inter-agency and Expert Group on SDG Indicators (IAEG-SDGs) into three tiers based on their level of methodological development and the availability of data at the global level.[2] However, this classification system does not assess the quality of the existing data on specific indicators, but only consider their conceptual clarity, the availability of an internationally established methodology and standards and the periodicity of data production at national level.

Data quality is recognized as a relevant issue for performance, both in relation to inter-organizational cooperation requirements (Batini & Scannapieco, 2006) and for decision-making activities (Chengalur-Smith et al., 1999). Improving the quality of existing data can help countries to identify data to assess SDGs performance and develop policies accordingly. The use of existing data is normally more cost-effective than the collection of new ones.

General criteria for evaluating the quality of data to benchmark SDG indicators are included in the United Nations National Quality Assurance Frameworks Manual for Official Statistics-UN NQAF Manual (United Nations, 2019).

The UN NQAF identifies the relevant principles to be assured while managing statistical process are:

1. Methodological soundness: statistical agencies should use sound statistical methodologies based on internationally agreed standards, guidelines or best practices in developing and producing statistics.
2. Cost-effectiveness: statistical agencies should assure that resources are effectively and efficiently used.
3. Managing the respondent burden: data collection should be balanced against production costs and the burden placed on respondents.

According to the UN NQAF, the relevant principles for ensuring the quality of statistical output are:

1. Relevance: Statistical information needs to satisfy user needs, considering the needs for global, regional, and national monitoring. Relevance is subjective and depends upon the varying needs of users (United Nations, 2019, p. 25).
2. Accuracy reflects the degree to which the information correctly describes the phenomena it was designed to measure, the degree of closeness of estimates to true values (United Nations, 2019, p. 8).
3. Reliability refers to the consistency of a measure over time and other conditions (Drost, 2011, p. 106), to the closeness of the initially estimated value(s) to the subsequent estimated value(s) (United Nations, 2019, p. 8).
4. Timeliness and punctuality: Timeliness refers to how quickly—after the reference date or the end of the reference period–the data and statistics are made available to users. Punctuality refers to whether data and statistics are delivered on the promised, advertised or announced dates.
5. Accessibility and clarity: statistics have to be found and obtained without difficulty (accessibility), be available and accessible to all users on an impartial and equal basis in various convenient formats in line with open data standards, and be presented clearly and in such a way that they can be understood and properly used. Provision should be made for allowing access to microdata for research purposes, in accordance with an established policy that ensures statistical confidentiality.
6. Coherence and comparability: statistics have to be consistent (coherent), meaning it should be possible to combine and make joint use of related data, including data from different sources. Statistics should be comparable over time and between geographic areas.

These principles can be used to assess the quality of existing national data on specific SDGs targets. However, few practical applications of these criteria to existing data collection have been developed (see for example UNESCO Institute for Statistics, 2017, p. 31 to evaluate national data on SDG 4–Education).

In order to address the scarcity of practical operationalization of these quality dimensions for the assessment of specific SDGs targets, this paper elaborates and operationalizes the above-mentioned principles to build a framework for corruption data quality (see Results).

In order to operationalize and contextualize the above-mentioned quality criteria, this paper uses the benchmarks included in the Manual on Corruption Surveys (UNODC-UNDP, 2018). This Manual was developed in order to support member states in generating baseline data and enhancing their data collection capacity, and represents the most recent and comprehensive guide for collecting primary corruption data developed by an international organization (UNODC-UNDP, 2018).

2.2 | Challenges in defining and measuring corruption

The reason for choosing corruption measurements as a reference for creating a practical framework on SDGs data quality is twofold.

First of all, fighting corruption is recognized as an important condition for the successful achievement of all the 17 SDGs (Rubio & Andvig, 2019). The 2030 Agenda recognizes corruption as an obstacle for sustainable development and devotes target 16.5, to "Substantially reduce corruption and bribery in all their forms".[3]

Secondly, the complexity of corruption and the lack of a standard and exhaustive definition (Heywood, 2017) inevitably affects the validity[4] of corruption metrics (Rose, 2018; Thatcher, 2010), and makes the assessment of corruption metrics an interesting case study for data quality evaluation more in general.

The most widely adopted definition of corruption is the one proposed by Transparency International (2013): "the abuse of entrusted power for private gain". While being synthetic and covering a broad range of corrupt activities, this definition has been increasingly criticized by academics. Criticisms mainly concern the normative and moral nature of this definition (Rose, 2018), the difficulty of identifying what constitutes an abuse (Heywood, 2017, p. 2), and the use of the term "gain" instead of "personal interest" (Marquette & Pfeiffer, 2015, p. 3). Furthermore, this definition is mainly based on a conception of corruption derived by the economic paradigm and principal-agent model and lacks an individual-level psychological lens considering the influence of power, self-control, loss aversion and risk acceptance (Dupuy & Neset, 2018). Another shortcoming related to existing definitions of corruption concerns the lack of attention to situational aspects that may influence corrupt behaviors (De Graaf, 2007, pp. 45–46). Another shortcoming of Transparency International's definition concerns the over-simplification of corruption and the lack of a clear understanding and definition of the specific behaviors constituting this phenomenon (Mugellini, 2020).

In order to address these issues, typologies have been often used to better distinguish corrupt behaviors not only on the basis of specific mechanisms, attributes, procedures and sectors, but also in relation to precise personal attitudes and motivations. The most common classifications of corruption are those distinguishing between grand and petty corruption (Rose-Ackerman, 1978), administrative and political corruption (Gould, 1991; Huberts, 1998; Navot, 2014, Holmes, 2015; OECD, 2015; World Bank, 2003), illegal and legal corruption (Kaufmann & Vicente, 2011; Maciel & De Sousa, 2018); the type of gain (tangible or intangible) involved in the transaction (UNODC, 2006; OECD, 2007; OECD, 2008; Villeneuve et al., 2019) and the principal's or agent's motivation for participating in the transaction (collusive vs. extortionary corruption) (Bauhr & Grimes, 2017; Jancsics, 2019; Ryvkin et al., 2017).

The complex nature of corruption inevitably affects also the measurement and policing of corrupt behaviors. Furthermore, as far as corruption is an illicit hidden behavior often characterized by some sort of collusion and co-responsibility of all parties in the corrupt transaction, it is rarely reported to competent authorities and/or discovered (Mugellini, 2020). As a consequence, administrative data on corruption, collected by criminal justice institutions (i.e., police, prosecution, courts) and anti-corruption authorities, suffer from a very high "dark figure" (the difference between the number of crimes experienced by a specific subject in a specific context and the number of officially recorded crimes). The lack of reliable data affects the possibility of inferential reasoning that is an essential process for understanding complex social phenomena.

Several attempts have been made to overcome the hidden nature of corruption. Perception-based metrics of corruption flourished in the 1990s with the aim of quantifying subjective opinions and perceptions of corruption among citizens, business representatives, civil servants or other stakeholders (e.g., Transparency International Corruption Perception Index). In the same period, indirect measures of corruption entailing a combination of subjective assessments, perceptions and other data that might be linked to corruption, also emerged (i.e., expert assessments and composite indices such as the World Bank Governance Indicators, the Global Integrity Index by Global Integrity and the Index of Public Integrity). These methods are mainly used for advocating for the fight against corruption, for awareness-raising activities guiding policy makers, investors and donors. Several scholars (Heywood & Rose, 2014; Knack, 2007; Mungiu-Pippidi & Dadašov, 2016; Sequeira, 2012; UNODC-UNDP, 2018) have highlighted validity, accuracy and explanatory shortcomings related to these methods. From a conceptual point of view, they were mainly criticized because they combine different types and shades of corruption, large number of variables referring to a wide spectrum of phenomena (Knack, 2007; Sequeira, 2012; UNODC-UNDP, 2018). Furthermore, perceptions-based corruption assessments appear to measure primarily administrative corruption rather than "state capture," or other corrupt behaviors, and appear to measure corruption in public procurement particularly poorly (Knack, 2007). Moreover, they often rely on individual believes and are sensitive to scandals, press reports, and political rhetoric but they rarely reflect the real level of corruption (Sequeira, 2012). For all these reasons they provide limited support for targeted evidence-based policy action (Knack, 2007; Sequeira, 2012).

Sample surveys have arisen as a reaction to these shortcomings. They can be used to collect data on both the perception and the experience of corruption on a representative sample of population (e.g., households, businesses, civil servants, etc.). They are based on solid and transparent methodology that is often standardized at international level, and use non-legal corruption definitions (Mugellini, 2020). For this reason, they provide data comparable across countries. Sample surveys also allow the gathering of micro-level data not only on corruption cases, but also on affected individuals (bribe payers and receivers) and their demographic, social, and economic background (Lynch, 2006; UNODC-UNDP, 2018). This information is fundamental to creating effective reforms and monitoring their efficiency. The main limits of sample surveys on corruption are linked to potential non-disclosure of information by the interviewees. However, this issues can be kept under control through a sound questionnaire design and wording. Sample surveys are also often accused to focus on the easiest concepts to measure, even if that means misinterpreting the prevalence and nature of corruption and focusing on corruption problems that are not the most relevant for that specific country or sector (Andersson, 2016; Chabova, 2017; Knack, 2007). In particular, sample surveys seem to mainly measure administrative corruption (Knack, 2007, p. 6). Even if business surveys

can provide information on some aspects of state capture (e.g., improper influence over laws and regulations affecting business) (Knack, 2007, p. 6), it is very unlikely that surveys will collect information on corrupt transactions occurring entirely within the state or on conflict of interest (World Bank, 2000). Chabova (2017) also highlights how public opinion surveys analyze only bribery, in many cases only passive bribery (i.e., whether the respondent has experienced request or hint for bribery) (Chabova, 2017, p. 1879). While bribery is easy to measure and to identify, it leads to a focus on low level corruption thus disregarding potential problems of systemic corruption. The focus on passive bribery only has also consequences on the choice of anti-corruption responses because being requested a bribe is a very different concept and entails different mechanisms and motivations than actually paying the bribe (Chabova, 2017; Villeneuve et al., 2019). The high costs for conducting sample surveys is also frequently mentioned among their shortcomings (UNODC-UNDP, 2018).

Despite these issues, sample surveys are widely recognized as the most appropriate tool for collecting corruption data and for complying with the methodological soundness requested for their estimation (UNODC-UNDP, 2018; Jandl, 2017[5]). Sample surveys are also considered the best method to collect data for building the two indicators, one for individuals (16.5.1[6]) and one for businesses (16.5.2[7]), identified by the UNStats for monitoring corruption under the umbrella of SDGs (Mugellini & Villeneuve, 2019; UNODC-UNDP, 2018; Jandl, 2017).

2.3 | Existing national and cross-national corruption surveys

The identification of surveys to test our framework on corruption data quality is partially based on the results of a desk review carried out in 2016–2017 within the context of an UNODC-UNDP project[8] that sought to determine the main corruption measurement tools developed at the national, regional and international level (Mugellini, 2018). This desk review has been completed and updated by the authors in order to cover surveys developed between 2005 and 2019. The search was conducted within electronic databases, journals and websites. In particular, online websites and databases of: (a) national statistical offices and other governmental agencies (e.g., Ministries of the Interior, Ministry of Justice, Anti-Corruption Commissions, etc.) from different countries; (b) international agencies (e.g., the World Bank, the European Bank for Reconstruction and Development, the World Economic Forum, the United Nations Office on Drugs and Crime, USAID); (c) international non-governmental organizations (e.g., Transparency International). The literature search was based on selected keywords (i.e., type of corruption; type of instrument; focus of the instrument; target population), searched separately and in combination. These keywords had been translated in French, German, Italian. and Spanish. Furthermore, the authors consulted a group of relevant scholars and international experts on corruption (i.e., members of the United Nations Office on Drugs and Crime Task force on Corruption Measurement, UNCAC Coalition, Transparency International, Open Government Partnership Initiative, and representatives of National Statistical Offices in countries with a long-standing experience in corruption surveys) in order to validate the completeness of the identified surveys' list.

The identification of both national and cross-national surveys is based on the following selection criteria:

1. *Main scope of the survey*: specialized surveys on corruption and other sample surveys (e.g., victimization surveys, governmental, environmental and social surveys, etc.) including a set of questions (at least two) on experiences with corruption are eligible for inclusion.
2. *Institutional/governmental nature of the responsible agency*: only national surveys developed by national institutional agencies (i.e., national statistical offices, or governmental agencies), and cross-national surveys carried out by international intergovernmental organizations (i.e., the United Nations, the World Bank, the Organization of American States, the European Commission), and international non-governmental organizations (i.e., Transparency International) are eligible for inclusion. The reason for this selection criterion is twofold: (a) as far as the selected surveys will be used to test and also "calibrate" our quality assessment framework, we prefer to focus on surveys that are usually considered methodologically accurate in order to ensure the best calibration. Academic surveys can be more valid in terms of content and problems conceptualization, but surveys developed by national statistical offices (NSOs) usually meet higher standards of methodological accuracy and reliability. Indeed, NSOs usually have adequate resources and infrastructures to address larger samples, use expensive data collection-methods such as face-to-face or CATI and access population and business registries for sample selection); (b) national institutional surveys are also more likely used by government for SDGs bench-marking. Indeed, national statistical systems play a leading role in the implementation and monitoring of the SDGs indicators framework (UNDP-UNODC, 2018). NSOs are officially involved and consulted by the custodian agencies[9] for compiling international statistics, producing regional and global aggregates and contributing to annual SDG progress reports, feeding into the follow-up and review processes (UNDP-UNODC, 2018, p. 19). The abovementioned international organizations are also relevant insofar as they carry out the majority of national surveys in less developed countries.
3. *Target population*: both population and business surveys are included.
4. *Update*: surveys developed between 2005 and 2019.
5. *Periodicity*: only cross-national surveys conducted at least twice are considered eligible. Periodicity was not used as selection criterion for national surveys.

On the basis of these selection criteria, 33 national surveys and 11 cross-national surveys have been identified. Table 1 below groups the identified surveys according to their focus (i.e., specialized on corruption and non-specialized), target population (population and business), and geographical scope (national or cross-national).[10] Specialized surveys on corruption are those specifically developed for measuring corruption issues, while non-specialized surveys are focused either on general safety/victimization issues or other social/governmental/

	National surveys		Cross-national surveys		
	Business	Population	Business	Population	Total
Specialized corruption surveys	5	9	1	2	17
Non-specialized surveys	4	15	2	6	27
Total	9	24	3	8	44
Total	33		11		44

TABLE 1 Summary of identified national and cross-national surveys including issues of corruption

environmental issues, and include a set of questions on corruption among other items. This distinction is relevant because specialized surveys on corruption are designed and structured to handle and overcome specific issues in collecting information on this complex phenomenon, and this might affect their level of validity and accuracy.

The desk review identified a majority of non-specialized surveys on corruption (27), among them there is a majority of victimization surveys on individuals and surveys on governmental issues with a block of corruption items (e.g., Governance, Democracy and Multiple Dimensions of Poverty [GPS-SHaSA], the Life in Transition Survey; the Afrobarometer and the Americas Barometer, etc.). Among the three non-specialized cross-national business surveys, the World Bank Enterprise Survey is the most updated and largest one. It covers 135 countries and it is often considered as a key-source of data for SDGs indicator 16.5.2, even if it collects data on gift and informal payment only with regard to tax officials.[11]

Seventeen of the surveys are specialized on corruption. The majority of them have been developed at national level, only three of them are cross-national projects: The Global Corruption Barometer (Transparency International), the Special Eurobarometer on Corruption and the Flash Eurobarometer on Corruption among businesses (European Commission). Six out of nine national population surveys specialized on corruption have been carried out by national statistical institutes in cooperation with the United Nations Office on Drugs and Crime. Similarly, nine out of 15 population non-specialized corruption surveys have been conducted under the supervision of the United Nations agencies. Notably across Latin American countries under the UNODC Latin American and the Caribbean Crime Victimization Survey Initiative, and across African countries under the Data for Africa initiative. Among national surveys on corruption, there are also projects "autonomously" developed by national agencies (i.e., in Chile; Italy, Luxemburg; Mexico; the Philippines; Zambia).

Relevant survey documents (e.g., questionnaires, methodological appendix, reports, etc.) were also reviewed for a comprehensive assessment of survey quality.

3 | RESULTS

3.1 | The quality evaluation framework for SDGs target 16.5

This section outlines the original framework developed by the authors for assessing existing corruption survey that can be used for monitoring SDGs target 16.5. The Quality Evaluation Framework for Sustainable Development Data on target 16.5 (QUEST_16.5[12]) (see Table 2 below) represents not only the first methodological steps of this study but also its first practical contribution and result.

The QUEST_16.5 mainly draws on the dimensions included in the United Nations National Quality Assurance Framework (UN NQAF) (United Nations, 2019) and operationalizes them (see section "*Existing criteria for assessing data quality of SDGs indicators*") by integrating elements of international goals for computing SDGs indicators on corruption, and methodological recommendations for measuring corruption (UNODC-UNDP Manual on Corruption Surveys). The UN NQAF has been chosen as a basis for the QUEST_16.5 because it is the most up-to-date framework on data quality developed by an international agency and it specifically addresses the quality of SDGs indicators.

The entire methodological process suggests the steps that should be considered to obtain ad-hoc quality assessment frameworks for other SDGs targets.

1. Relevance refers to the ability of surveys' results (statistical output) to satisfy the needs of the user (United Nations, 2019, p. 8) and, in particular, the needs for global, regional and national SDGs monitoring. For the purpose of this paper, relevance is determined by evaluating whether the survey provides data useful to monitor corruption development according to Target 16.5. SDGs indicators on corruption suggest monitoring the level of corruption by (a) including a screening question to detect those persons/business who had contact with public officials in the previous 12 months, (b) measuring bribes paid/given to public officials, and (c) bribes requested by public officials but not paid, in the past 12 months.[13] Ideally, all three elements (a, b, and c) should be covered by the survey to produce relevant data for target 16.5. However, even surveys collecting information only on (a) and (b) can provide a reasonably close approximation of the prevalence rate of bribery (Jandl, 2017, p. 118). Therefore, the relevance of existing surveys' data has been evaluated as the ability of surveys to provide information on (a) and (b) for each SDGs indicator on corruption. Also the data disaggregation by type and sex of the public official has been considered as relevant for SDGs. This said, it is important to mention that SDGs indicators on corruption refer only to one specific form of corruption–bribery, and this might be considered as a limitation. Indeed, as far as forms of corruption are theoretically expected to vary across countries (as well as within them), treating corruption as one-dimensional phenomenon and using bribery as a suitable proxy of corruption

TABLE 2 The quality evaluation framework for sustainable development data on target 16.5 (QUEST_16.5)

Area	Quality dimension	Sub-dimension	Operationalization	Categories of the dimension
Statistical output	1. Relevance	SDGs indicators on corruption	Inclusion of questions on: (a) contact with public officials; (b) bribes paid/ given to public officials; (c) bribes requested by public officials but not paid	Yes = coverage of at least (a) and (b)no = lack of coverage of either (a) or (b), or both of them
Statistical process (assuring methodological soundness) and statistical output	2. Accuracy/validity	N. Items covered	Coverage of UNODC-UNDP recommendations	For pop. Surveys: High (18–13 topics); medium (12–6); low (less than 6)for business surveys: High (21–14 topics); medium (13–7); low (less than 7)
		Question formulation	Coverage of UNODC-UNDP recommendations)	High (3 out of 3); medium (2 out of 3); low (1 out of 3)
	3. Reliability	UNODC-UNDP methodological recommendations	Recommendations on target population; sampling unit and respondents selection	Pop. Surveys: High (3 out of 3): Medium (2 out of 3); low (1 out of 3)bus. Surveys: High (4/4 out of 4); medium (2 out of 4); low (1 out of 4)
		UNODC-UNDP methodological suggestions	Suggestions on survey mode, pilot survey, sample design; sample size	Pop. Surveys: High (5/4 out of 5): Medium (3/2 out of 5); low (1 out of 5) Bus. Surveys: High (4/3 out of 4); medium (2 out of 4); low (1 out of 4)
Statistical process and output	4. Periodicity	—	N. of times the survey has been done	High (more than twice), medium (twice), low (once)
Statistical output	5. Accessibility	Access to survey documents	Publicly accessible questionnaire; methodology; executive summary/introduction to tool; reports with main results	High (questionnaire, methodology and analysis of results available)medium (no full questionnaire available but reports/presentations with main questions and methodology)low (no full questionnaire or methodological details)
		Access to raw data	Publicly accessible raw data	Yes; no
Statistical process and output	6. Comparability	Standards for questions	Coverage of: (1) SDGs 16.5 key elements; and (2) UNODC-UNDP recommendations on questions formulation	High (at least two of the three elements of SDGs 16.5 and all 3 recomm.); medium (at least two of the three elements of SDGs 16.5 and at least 2 recomm.); low (none of the three elements of SDGs 16.5 and less than 3 of the recomm.)
		Methodology standards	Average of UNODC-UNDP score for recommendations and suggestions	High (high score for recomm. And sugg.); medium (medium score for recomm. And sugg.); low (low score for recomm. Or sugg.)

across countries might lead to "misinterpretation of the prevalence and nature of corruption in the given setting, and to a focus on corruption types/problems that are not the most pressing" (Andersson, 2016, p. 69). For this reason, the paper also assesses the validity of existing surveys by checking the inclusion of questions that might help to detect other forms of corruption than bribery (e.g., favoritism, vote buying, etc.).

2. Accuracy/Validity refer to the "closeness of estimates to the exact or true values that the statistics were intended to measure" (United Nations, 2019, p. 8) According to Drost (2011, p. 115) there are four types of validity: statistical conclusion validity, internal validity, construct validity and external validity. For the purpose of this paper, we mainly consider construct validity[14] that refers to how well a concept, idea or behavior is operationalized, translated

into a functioning and operating reality (Drost, 2011, p. 115). A valid operationalization process ensures that a specific survey measures what the researcher intended to measure. We determine construct validity of existing surveys by evaluating the compliance with UNODC-UNDP Manual's recommendations on the number and types of questions on the experience of corruption and on how to phrase them (see Table S1 in the Supporting Information for further details).Besides the number and types of information collected on a given phenomenon, the construct validity of a survey-based measure is heavily dependent on the presence of a detailed operational definition of the phenomenon under investigation. A clear operational definition of corruption, containing all the relevant attributes that define the particular behavior, and using practical examples, induces respondents to search their memories for events with all those attributes (Lynch, 2006) and avoids them making personal interpretations of the phenomenon at question (Mugellini, 2018). A precise wording can facilitate the disclosure of an accurate response and minimize memory decay and social desirability bias (UNODC-UNDP, 2018). The UNODC-UNDP Manual on Corruption Survey, 2018, (p. 85) suggests to avoid the generic word "bribery", because in many cases respondents do not consider their experience as a form of bribery, and instead refer to the exchange of money, goods or favors.

3. Reliability refers to the consistency, stability, of the measurement over time and other conditions (e.g., populations) For the purpose of this paper, the reliability of existing surveys is determined by assessing the compliance with international methodological recommendations and suggestions on data collection methods and sampling design (UNODC-UNDP, 2018). In particular, for population surveys, the compliance with the following requirements have been checked: (a) the target population refers to persons aged 18 years or older[15]; (b) the sampling unit refers to individuals and not households (UNODC-UNDP, 2018, p. 72); (c) one respondent is randomly selected in the household.[16] With regard to recommendations for business surveys, the compliance with the following requirements have been checked: (a) the target businesses belong to specific economic sectors and size (see Table S1 in the Supporting Information for further details); (b) the sampling units are individual business establishments (instead of business entity); (c) the type of respondent is appropriate to the size of the business (see Table S1 in the Supporting Information for further details).
   In addition, it has been investigated whether (a) the survey mode is based on face-to-face and CATI interviews[17]; (b) cognitive testing and or pilot surveys have been conducted prior to the full survey; (c) the survey sample is stratified by geographical conglomerates (and sector, n. employees for business surveys); (d) the size of the sample guarantees representatives results by main geographical conglomerates.
4. Periodicity refers to the frequency of a measurement. For the purpose of this paper, periodicity considers the number of waves of surveys. Even if not specifically indicated in the UN NQAF, the authors consider the periodicity of the survey as a relevant dimension of quality. It indicates the continuity of the survey (important for the statistical process), and the availability of recurrent data (important for the statistical output) that can be used for policy design, policy evaluation, and policy learning.
5. Accessibility refers to the ease with which statistical information can be obtained (United Nations, 2019, p. 8). For the purpose of this paper, accessibility evaluates the availability of specific publicly accessible documents on the surveys (i.e., Questionnaire; Methodology; Executive Summary/Introduction to tool; Reports with main results), and raw data. Data access is essential for governments to measure the progress towards the Agenda 2030 and SDGs, and for the civil society and the public to track progress.[18]
6. Comparability refers to the extent to which differences in statistics from different geographical areas, non-geographical domains, or over time, can be attributed to differences between the true values of the statistics (United Nations, 2019, p. 8) and not to differences in the data collection methods or concepts operationalization. For the purpose of this paper comparability of surveys' results is determined by evaluating whether the surveys follow international recommendations on how to structure corruption questions (coverage of SDGs 16.5 key elements and UNODC-UNDP recommendations) and on the data collection method (average of the score obtained for the UNODC-UNDP methodological recommendations and suggestions for reliability). This ensures that results are comparable over time and between geographical areas. Comparability partially overlaps with the dimension of relevance, validity and reliability.

The analysis does not consider the dimensions of timeliness and punctuality mentioned in the UN NQAF (United Nations, 2019, p. 8) because it is generally difficult to determine when exactly surveys' data was published or released by the responsible agency. Also the coherence of surveys' results across datasets, over time and across countries is not considered in this analysis because it would require an ad-hoc research and detailed quantitative analyses of surveys' results.

3.2 | Testing the QUEST 16.5

Having carefully designed the assessment framework, it will be used to test the previously identified national and cross-national corruption surveys. They are evaluated with regard to their content, methodology, and results based on the QUEST_16.5 dimensions. Different colors indicate different levels of quality for each dimension: high (green), medium (yellow), and low (red). Blank cells indicates the lack of data to evaluate that specific quality dimension. The overall quality score is a synthetic indicator computed as an average of the scores on each of the 10 quality sub-dimensions. It ranges from 3 (maximum level of quality) to 0 (minimum level of quality). The following tables include national surveys evaluation (Table 3) and cross-national surveys assessment (Table 4).

TABLE 3 Quality assessment of existing national surveys on corruption [Colour table can be viewed at wileyonlinelibrary.com]

	Accuracy/validity			Reliability		Accessibility			Comparability		
Acronym of national surveys	Relevance	Items covered	Quest. Formulat.	Method Recomm.	Method Suggest	Periodicity	Survey documents	Microdata	Standard questions	Standard method.	Quality score
Specialized corruption surveys											
ACBS_P_IND											N.A
SCPP_B_ITA											1.5
SEC_P_PHL											1.9
ENCIG_P_MEX											2.6
SGNS_P_ZMB											1.3
CS_P_WBs											2.6
CS_B_WBs											2.5
CS_P_AFG											2.5
CIS_P_IRQ											2.3
CER_P_NGA											2.5
CS_B_NGA											2.4
NSCE_P_KEN											2.3
NSCE_B_KEN											2.4
CFI_B_ETH											2.1
Non-specialized surveys											
NSPS_P_CHL											1.9
SC_P_ITA											2.5
V&S_P_LUX											1.4
ENVE_B_MEX											2.3
ENVS_P_VEN											1.7
ENV_P_ARG											1.7
ENPEVI_P_GTM											2.0
ENVI_P_PAN											2.5
VS_P_KEN											1.2
VS_P_EGY											1.2
VS_P_TZA											1.2
VS_P_UGA											1.2
VS_P_RWA											1.2
SCC_P&B_CPV											1.3
VOCS_P_ZAF											1.5
NCVS_P_JAM											1.9
BCS_B_ITA											2.0
ENCRIGE_B_MEX											2.1

Note: Supporting Information (Table S2) for the complete surveys' names. Different colors indicate different levels of quality for each dimension: high (green), medium (yellow), and low (red).

TABLE 4 Data quality assessment of existing cross-national surveys on corruption [Colour table can be viewed at wileyonlinelibrary.com]

		Accuracy/validity		Reliability			Accessibility		Comparability		
Acronym of cross-national surveys	Relevance	Items covered	Quest. Formulat.	Method Recomm.	Method Suggest	Periodicity	Survey documents	Microdata	Standard questions	Standard method	Quality score
Specialized corruption surveys											
EBS_P_EC											2.8
EBF_B_EC											2.7
GCB_P_TI											2.3
Non-specialized surveys											
ICVS_P_UN											2.1
ES_B_WB											2.6
BEEPS_B_EBRD-WB											2.5
LiTS III_P_EBRD											2.6
AFRBAR_P_AFR&TI											2.4
AMEBAR_P_LAPOP											2.4
GPS_SHaSA_P											2.4
EQI_P_QoG_EC											2.5

Note: Supporting Information (Table S3) for the complete surveys' names. Different colors indicate different levels of quality for each dimension: high (green), medium (yellow), and low (red).

3.3 | Quality assessment of existing national surveys on corruption

3.3.1 | Relevance

Twelve out of 24 identified national population surveys provide data that are relevant for monitoring SDGs target 16.5. The same result is registered for national business surveys, where four out of nine are able to produce data for computing SDGs indicators on corruption. All identified specialized corruption surveys are able to provide data for monitoring SDGs target 16.5, except the State of Governance National Survey in Zambia that does not include the screening question on the contacts with public officials. Only four out of 18 non-specialized surveys provide information for SDGs indicators on corruption. The main issue with this type of surveys is the lack of the screening question on whether respondents had (or not) contacts with public officials. The lack of this question significantly affects the comparability of surveys' results because the percentage of bribes will be different when computed as a ratio on the total number of interviewees, or as a ratio on those who had a contact with a public official.

3.3.2 | Accuracy/validity

The main validity issue of corruption surveys is linked to the fact that the majority of them cover only bribery and do not include other, more complex, corruption types. There are a few exceptions: The Italian Citizens Safety Survey includes questions on vote buying and favoritism; the corruption surveys in Ethiopia ask about favoritism and nepotism, the Mexican Survey of Quality and Government Information covers the misappropriation or other diversion of property by a public official.

In addition, the majority of identified national surveys cover only passive bribery,[19] and neglect those cases when the citizen actively offers a bribe to public officials. The main reason for this limited focus of corruption surveys is to avoid non-responses due to social stigma and social desirability, which is observed when respondents are asked about acts of active bribery. Being requested a bribe entails different practical and psychological mechanisms than actively offering a bribe. Respondents, who feel they were obliged to pay a bribe, feel less guilty or responsible, and are thus more likely to provide an accurate answer. However, focusing on passive bribery only leads to a mis- and underrepresentation of corruption.

The Manual on Corruption Survey (UNODC-UNDP, 2018, p. 85) suggests to use a wider formulation of the main question on corruption experience to include also those cases in which the public officials did not ask for the bribe but the counterpart offered it (i.e., "Did you have to give extra money, gift, a favor, etc."). Some of the most recent national surveys entail a similar strategy (e.g., the Corruption survey in the Western Balkans, in Afghanistan, in Nigeria, asks the following: "Did it happen that you had to give to any of them a gift, a counterfavor or some extra-money…"), while other surveys include a very direct question on active bribery (e.g., the Corruption Surveys in Nigeria 2016, 2019 ask: "During the last 12 months, was there any occasion where you offered, directly or indirectly, to give extra money or a gift to a public official (in addition to the correct amount of official fees) for an issue or procedure related to his/her function but the public official refused the offer?) in addition to passive bribery.

Furthermore, the majority of identified surveys cover only bribes paid/given to public officials but do not consider bribes requested but not given. Despite the refusal of providing the bribe, such events are also classed as bribery, as the UNCAC definition of bribery states (UNODC-UNDP, 2018, p. 85). The lack of this information can affect the validity of indicators 16.5.1[20] and 16.5.2[21] for SDGs monitoring. Indeed, it impedes comprehensive measurement of the phenomenon under investigation because it focuses only on those bribe requests that have been paid. Among the identified national surveys, only seven out of 33 include questions able to ascertain whether the requested money, gift or favor was provided or not (i.e., the population and business corruption surveys in Nigeria; the Italian Survey on Citizens' Safety, the population and business surveys in the Western Balkans; the Mexican Survey of Quality and Government Information; the National Household Survey on Experience with Corruption in the Philippines).

Data disaggregation is also relevant for the validity of survey data because, by providing more details on the phenomenon, it contributes to its better understanding and allow for more effective policies. Eight out of 16 specialized corruption surveys disaggregate the prevalence of bribery by type of public officials. However, this information is missing in the majority of non-specialized surveys. They either collect an aggregated figure or they distinguish by type of service instead than by type of public officials.[22] Five out of nine national business surveys are able to provide information on the prevalence of bribery by type of transaction/procedure. The majority of those are non-specialized corruption surveys. Even if the majority of identified surveys measure corruption experience by type of public officials, or private entities, only few include follow-up questions on the types of procedure during which corruption was expected, modus-operandi (direct request or through an intermediary), type of requested benefit and value of the benefit. The lack of this information affects validity because it does not allow for the distinction between petty and grand corruption, administrative and political corruption, favoritism and bribery. All surveys include a question on the frequency of corruption requests, highlighting the importance of understanding the recurrence of this phenomenon and distinguishing between sporadic cases and more "organized" mechanisms.

Specialized surveys provide a detailed operational definition of corruption, while the majority of non-specialized surveys (both national and cross-national) fail in the formulation of the question on the experience of bribery. Indeed, they do not describe the corruption event under investigation but simply mention the term "bribe", "bribery" or "corruption". In this way, they run the risk of fostering social desirability bias. Business surveys show a higher compliance with international benchmarks related to the

formulation of the question on the experience of corruption than population surveys.

3.3.3 | Reliability

The methodological accuracy of selected surveys is generally high. The main problems concern the characteristics of the target population and the way respondents are selected. Indeed, the majority of victimization surveys focus on individuals aged 14 and older while the recommendation for collecting accurate results on corruption among individuals is to target people older than 18 who, in the majority of countries, are legally considered to be adults and can interact with public authorities. Incoherent respondents' age also affects the comparability of survey's results.

In addition, some of the identified surveys use households as sampling unit, instead of individuals, and this can cause underestimation of corruption prevalence and incidence, as the household member selected as the respondent may not have complete information about all contacts with public officials and relevant bribe payments (UNODC-UNDP, 2018, p. 72).In some cases, respondents are not randomly selected within a household, and this can cause clustering effects, together with unnecessary burden for respondents leading to inaccurate responses (UNODC-UNDP, 2018, p. 75).

With regard to methodological suggestions,[23] the majority of national population surveys employ face-to-face interviews that ensures high response rate and accuracy. All surveys are based on large samples guaranteeing representatives results by main geographical conglomerates (and sector and number of employees for business surveys). Information on the development of pilot surveys and cognitive testing was not available in many cases.

3.3.4 | Periodicity

The majority of identified national surveys (18) have been carried out only once between 2005 and 2019. However, seven specialized corruption surveys and six victimization surveys with a module on corruption have been developed more than once.

3.3.5 | Accessibility

Relevant survey's documents (questionnaire, methodology and analysis of results) are completely accessible for 16 out of 44 identified surveys, more or less equally distributed across specialized and non-specialized surveys. For the rest of identified surveys there is no questionnaire available online, but reports/presentations with main questions and methodology. This issue can negatively affect the interpretability and use of the data by external users. The accessibility of raw data is low for almost all identified survey, with the exception of the surveys developed by the National Institute of Statistics and Geography of Mexico (INEGI) that publishes raw data of several surveys in different formats.

3.3.6 | Comparability

Only nine of the 32 identified national surveys show a high level of comparability with regard to the content of the surveys (i.e., use of international standard for the questions on the experience of corruption). The level of comparability, in terms of questions covered, is higher for specialized corruption surveys than for non-specialized ones. Indeed, many of the considered national victimization surveys[24] follow international standards for victimization surveys and are actually comparable with the International Crime Victims Surveys. However, they still do not comply with international standards on the measurement of corruption because they do not include a screening question on contacts with public officials. When evaluating comparability level, it is also important to consider the coverage of different corruption and bribery types (active and passive). For example, comparing the results of surveys including both active and passive bribery with the results of surveys covering only passive bribery would be misleading in terms of bribery prevalence.

3.4 | Quality assessment of existing cross-national surveys on corruption

On average, cross-national surveys show better results on almost all dimensions of quality with the exception of validity. The methodological quality of cross-national surveys is usually high but their ability of measuring different facets of corruption, or simply different bribery types, is weaker than national surveys. This issue affects the possibility of using results for policy-making.

3.4.1 | Relevance

All cross-national surveys collect relevant data for SDGs indicators on corruption, except the International Victimization Survey because it does not include a question on the contacts with public officials.

3.4.2 | Accuracy/validity

Identified cross-national surveys show a low level of validity both in relation to the questions' formulation and in terms of items covered. Indeed, the majority of them mention the words “bribe”, “bribery” or “corruption” in the main question on the experience of corruption, instead of providing a more “neutral” and operational definition of the phenomenon under investigation. The Special Eurobarometer on Corruption complies with these recommendations and define corruption as “offering, giving, requesting and accepting money, gifts and

favors to/for a public official in order to obtain a specific service, perform a procedure or avoid sanctions" (European Commission, 2017). Similarly, the Life in Transition Survey, the Enterprise Survey, and the Business Environment and Enterprise Performance Survey do not directly mention sensitive words (i.e., bribery, corruption, etc.) in the formulation on corruption experience. None of the cross-national surveys is able to cover all recommended items for the measurement of corruption. The lower number of questions is mainly due to the need of ensuring the sustainability of the survey across different countries, reduce respondents' burden and guarantee consistent response rate. As mentioned for national surveys, also the majority of cross-national surveys cover only passive bribery (when the public official or private entity asks, solicits, or lets the citizen understand a bribe is necessary to obtain a specific service) (Chabova, 2017, p 1879). For example, the Special Eurobarometer on Corruption targeted at individuals[25] and businesses[26] and the United Nations International Crime Victims Survey.[27] While the Global Corruption Barometer should be able to capture both passive and active bribery as it asks "Have you paid a bribe to any one of six public services in the past 12 months?" (Transparency International, 2013, p. 31).

3.4.3 | Reliability

The level of methodological accuracy is high for the majority of cross-national surveys. Minor issues are mainly related to the selection of business entities instead of individual business establishments as ultimate sampling unit for business surveys.

3.4.4 | Periodicity

All but one cross-national surveys have been carried out more than twice during our reference period. This result highlights their solidity in terms of methodology and the possibility of using their data to monitor corruption over time.

3.4.5 | Accessibility

The level of accessibility of surveys' documents is high for all cross-national surveys. However, microdata for five out of 11 surveys are still not publicly accessible.

3.4.6 | Comparability

The level of comparability is high for the majority of cross-national surveys. This result can be explained by the fact that these surveys need to be developed in several countries and are coordinated by those international agencies that are responsible for setting comparability standards.

4 | DISCUSSION

The paper identifies 17 specialized surveys on corruption, and 27 surveys including a set of questions on corruption in a more general questionnaire. Results show a higher level of quality of specialized surveys than general surveys with corruption-related modules, both for national and cross-national projects. The analysis identified specific issues that limit the quality of existing data sources and more generic challenges in measuring corruption. Some of the identified issues represent minor deviances but do not generate problems beyond the specific confines of the survey. Other aspects raise fundamental questions for the understanding of target 16.5 and allow to identify systemic challenges and blockages in the generation and collection of quality data for SGDs and policymaking as a whole (e.g., international political value of indicators rather than local usefulness).

4.1 | Different quality and use of cross-national and national surveys

On average, cross-national surveys show better results on almost every dimension of quality, with the exception of accuracy/validity. In particular, cross-national surveys show more difficulties in disentangling and covering different corruption and bribery types than their national counterparts. They make corruption issues more comprehensible to target audiences and, by comparing and ranking countries on the extent of corruption, they attract attention and encourage action. Thus, cross-national data seem to be more suitable for advocacy and comparability purposes across countries and over time rather than for policy-making. They can also be used, for example, to identify effects of international recommendations and legislations, to understand the role and effect of international and national anti-corruption bodies, and on specific national anti-corruption regimes (e.g., open data in public procurement).

National surveys, in contrast, are by definition focused on national/local corruption issues. As a consequence, the range of corruption types included in national surveys is usually wider capturing the national context better than cross-national surveys. For example, the Italian population corruption survey covers not only bribery but also vote buying and favoritism and includes a detailed set of questions on corruption in the health sector. Equally, the Italian survey on businesses is mainly targeted at corruption in public procurement. The focus of these surveys clearly highlight that in Italy there is a public interest in collecting data for monitoring corruption issues in three main sectors: health; public procurement and elections. The Mexican survey uses a very specific definition of corruption and refers to direct bribery requested by official, and bribery requested through an intermediary. In Mexico, as well as in other Latin American countries, informal intermediaries ("tramitadores" or "coyotes") frequently assist individuals and firms with procedures at the government bureaucracy (Fredriksson, 2014). Scientific evidence on this issue is extremely relevant when planning anti-corruption policies that, in order to be

effective, should target also these actors. Moreover, national surveys usually entail larger samples than cross-national ones allowing for analyses that are representatives at a micro-level (e.g., cities, or even municipalities). For all these reasons, national surveys are able to provide information that might be useful for informing national but also local policies, for identifying corruption-prone areas, procedures or positions at risk, population groups more exposed to corruption; or monitoring trends over time.

4.2 | Internationalization of data collection efforts

The low number of national corruption surveys is explained first through the fact that the analysis is focused only on surveys developed by national statistical offices or other governmental agencies, while surveys developed either by private or academic institutes were not included in this analysis. Second, it appears that national governments often rely on corruption data collected international organizations. This might indicate an internationalization of data collection efforts, or that national governments focus more on the political value of corruption measurement and compliance with international requirements, than on the need of collecting policy-relevant evidence. These results raise issues of reliance on external actors for generating fundamental evidence, as well as concerns regarding context specificity in gathering data. Cross-national surveys operationalize corruption in a way that allows to obtain results that, while compliant with international standards and comparable across-countries, fail to reflect local corruption problems and specificities. The mid-term risk of this approach is an ill-fitting harmonization of corruption data, a flattening of corruption problems that might lead to a loss of crucial information regarding local and cultural dynamics. In the long-term, this approach might also impact domestic anti-corruption efforts aligning them with international norms and policy ideas but not focusing on context-specific, local, corruption issues. This might in turn lead to the failure of anti-corruption policies.

Nevertheless, as noticed by Chabova (2017), in recent years, measuring corruption has been emerging as a priority also for national governments. This might indicate that national governments are becoming more and more aware of the need of developing national corruption surveys that comply with international recommendations but that are also targeted at contextual and local issues.

4.2.1 | Defining the population

The main issue related to the relevance of national surveys' data for SDG Target 16.5 is the lack of a screening question identifying the respondents who had a contact with a public official (or private entity) to obtain a specific service or perform a specific procedure during the reference period of the survey. This question allows to identify respondents that have been exposed to the risk of corruption, and helps them to facilitate their recollection of relevant events. The lack of this question might lead to less specific and accurate answers. This issue affects not only the relevance and validity of surveys' results but also their comparability. The percentage of bribes will be different when computed as a ratio on the total number of interviewees, or as a ratio on those who had a contact with a public official. If most people do not interact with public organizations, should we be satisfied with the fact that they have not witnessed bribery/corruption cases? Should they simply be removed from the equation? The screening question might make numbers look worse (higher levels of corruption), but be more accurate and more actionable in the end. The introduction of screening questions will not increase neither survey cost nor the burden for respondents while greatly improving the relevance of surveys' data for SDGs and their general validity and comparability.

4.2.2 | Measuring what is easiest to measure

The results confirm a common concern in the literature (Chabova, 2017; Knack, 2007; World Bank, 2000), namely that the majority of surveys focus on the easiest concepts to measure, that is, passive bribery. Furthermore, several surveys cover only bribes paid to public officials and neglect all those requests that have not been granted. Only the most recent surveys seem to address this challenge, covering different forms of corruption.

4.2.3 | Social desirability issues

Another validity issue, affecting both national and cross-national surveys, concerns the use of the term "bribe", "bribery" or "corruption" in the main question on the experience of corruption, instead of providing a more "neutral" and operational definition referring to the exchange of money, goods, favors, or services. This might lead to social desirability bias and underestimation of the prevalence of corruption. Indeed, the terms bribery and corruption are conventionally considered as undesirable acts and respondents might be less likely to report them when they are explicitly mentioned. Avoiding the term bribery is also preferable because in many cases respondents do not consider their experience as a form of bribery (UNODC-UNDP, 2018, p. 85). This observation touches some of the core challenges of measuring complex social phenomena and highlights the need of avoiding terms that might be culturally biased. Amongst the surveys considered in this paper, many remain insensitive to the social desirability issue associated with specific wording.

4.2.4 | Data richness

The variety of questions on the experience of corruption are fundamental to understand the precise circumstances of corrupt events. They enable the production of actionable information that can be directly used for evidence-based policies targeted at, for example, procedures that are particularly vulnerable to bribery, competent agencies, relevant regulations and costs. However, only few of the

identified surveys include follow-up questions on the types of public officials involved, procedures during which corruption was expected, modus-operandi (direct request or through an intermediary), type of requested benefit and value of the benefit. The lack of this information does not allow for the distinction between petty and grand corruption, administrative and political corruption, favoritism and bribery.

4.2.5 | Data disaggregation

The possibility, in most surveys, of disaggregating the data by socio-demographic characteristics of respondents is in line with international recommendations. Besides providing important data for orienting anti-corruption policies, disaggregated information also answers the SDGs principle of leaving no one behind.

4.2.6 | Methodological soundness

The methodological accuracy of selected surveys is generally high. Indeed, the experience and infrastructures of national statistical authorities and international organizations in the survey design and data collection phase represent a guarantee for data quality. The main methodological issues concern the characteristics of the target population and the way respondents are selected, but they are easily solvable.

4.2.7 | Periodicity of national surveys

The majority of selected surveys have been conducted only once. This is problematic, since "the value of surveys will be much enhanced by repeatedly carrying out comparable waves that allow the measurement of trends over time, the evaluation of anti-corruption measures and prompt corrective action to be taken" (UNODC-UNDP, 2018, p. 34). The suggested approach, in view of the evolving and constantly changing nature of corruption, is to carry out corruption surveys at least every 3 years to meet the requirements of monitoring corruption trends within the SDGs framework (UNODC-UNDP, 2018). This would lead to better and more consistent data, but also allow for greater policy learning. This calls for a more consistent economic support from national and international agencies to avoid that a relevant know-how is wasted because of the scarcity of resources.

4.2.8 | Data accessibility

For both national and cross-national surveys, the difficulty in accessing the surveys' raw data emerges. This is particularly problematic considering that one of the most praised advantage of sample surveys is to provide data at the highest level of disaggregation. This also runs against the transversal dynamics implied in the SGDs process, which is openness and transparency, notably through open data initiatives. Furthermore, the release of microdata increases transparency and thus serves to promote trust in surveys' results (UNODC-UNDP, 2018, p. 27). Indeed, it addresses the possibility of manipulation of results in order for states to appear more performing than they are in regards to SDGs but also other international evaluative frameworks.

Overall, identified quality challenges mainly concern the difficulties in translating complex phenomena in functioning and measurable indicators. The selection of the most easily operationalizable aspects will lead to clear results, but at the cost of relevance. The impact will then be in the development of truncated and incomplete policy options that might miss out the most salient and important aspects of the problem, specifically because they are harder to measure. However, the paper demonstrates that these challenges can be mitigated by defining what is needed to measure, being clear what the data is used for and choosing survey methodology appropriate, integrating good international standards, but also looking out for examples on how to tackle local persisting problems. There is a road ahead for measuring complex phenomena present across all 17 SDGs and obtaining quality data, but this road needs to be paved with clear international benchmarks, local practical operationalization and periodical quality assessments.

5 | CONCLUSIONS

This paper represents an original quality assessment of baseline data for measuring target 16.5 on corruption. It not only identifies existing data sources for measuring corruption but also develops a dedicated quality assessment framework (QUEST_16.5) to evaluate their quality. Results point both to the best-in-class efforts and to the most common informational and methodological gaps that need to be addressed. This is crucial not only to develop solid and sustainable monitoring systems for SDGs but also to make the best use of available statistical information for measuring corruption.

The proposed framework identifies the international benchmarks that are necessary for an effective quality evaluation of corruption metrics. The contextualization of generic quality dimensions and their operationalization according to specific benchmarks represents a starting point that can be replicated in order to test the quality of existing metrics related to other SDGs indicators.

Results are useful for policy-makers because they identify the current shortcomings of existing national official surveys, beyond SDGs requirements, and provide suggestions on how to overcome them by looking at international guidelines but also considering national priorities. Data quality should never be taken for granted even for data that are produced by national statistical offices or international agencies. In this regard, QUEST_16.5 can be used by representatives of national statistical offices (NSOs) to check and improve the quality of national corruption surveys and measurements.

Raising awareness on the importance of data quality assessment is even more relevant considering that many countries are enthusiastic about partnerships with private providers of big data for answering

the demand for statistics on SDGs (Merry, 2019). These new sources of data "seem to solve the resource problem for poor countries, but raise issues of transparency, accountability, and the possibility of further weakening national statistical offices" (Merry, 2019, p. 147). The key role of NSOs in gathering SDGs data and their responsibility for validating data, adopting internationally agreed standards and assuring quality has been widely recognized (Merry, 2019; UNStats, 2017). Statistical agencies, and international organizations should consider among their tasks the accreditation and certification of data sets created by third parties or private sectors (MacFeely, 2019, p. 126). In particular they should keep "control of quality and limit the risk of private big data producers and users fabricating data sets that fail the tests of transparency, proper quality, and sound methodology" (Hammer et al., 2017, p.19). The QUEST_16.5 could support this additional task of NSOs and international organizations and be used to check the quality of third parties and private sector data in measuring corruption. Similarly, big data producers, such as non-governmental organizations (e.g., NGOs and universities), can use QUEST_16.5 to develop a "self-assessment" of the quality of their corruption metrics.

For academics, this paper facilitates secondary data use by analyzing the content of existing national official surveys on corruption and identifying their level of validity, accuracy and comparability. It also supports the ex-post harmonization of survey data necessary to compare surveys results across countries. Furthermore, it provides a repository of information on existing corruption surveys and raises awareness among potential users of the existing data richness.

The main limitations of this analysis is the exclusion of those surveys developed by private research institutes and organizations, and those surveys for which documents are not publicly available, or published in a language different from those covered in the literature search.

Future research should better test the validity of survey's outputs by using specific quantitative tests (e.g., test–retest reliability, alternative forms, split-halves, inter-rater reliability, and internal consistency) and their coherence over time and across countries. The same analysis could be also replicated for corruption surveys developed by private organizations. This would represent an additional quest for identifying high-quality indicators for SDGs monitoring, but also to question and improve the current set of SDGs indicators in order to make it more suitable for national and local policy design, and not only for political purposes.

Indeed, a common problem to all SDGs indicators is that they cover only small parts of each goal (Merry, 2019). This paper confirms this concern demonstrating how SDGs indicators on corruption mainly focus on the easiest concept to be measured–bribery–instead of addressing corruption in all its forms. Besides validity problems, this approach might also raise political issues. Indeed, it has been demonstrated that bribery is a type of corruption that occurs more frequently in developing than developed countries, where corruption problems are mainly linked to grand corruption (Graycar, 2015; Johnston, 2005; Johnston, 2013). The SDGs approach to corruption measurement, somewhat neglecting grand corruption, might underestimate corruption problems in more developed contexts, and thus fail to properly monitor corruption trends in these environments.

Furthermore, many of the existing SDGs indicators provide evidence of problems but do not address those factors that might help to solve the problem (Merry, 2019). In the case of corruption, they provide information on the prevalence of bribery in a given country but not on the presence of specific anti-corruption policies and their effectiveness.

Finally, this analysis confirms the political value of quantifying corruption and its influence on state behavior and global governance (Cooley & Snyder, 2015). It demonstrates that not only the results of corruption metrics, but also the production of corruption measurements are highly politicized (Wickberg & Mugellini, 2020). Corruption metrics and rankings often form the basis for foreign investment decision, blacklisting of countries or significant processes such as countries' accession to the European Union. At the same time, the development of corruption measures is frequently seen as a mere necessity for meeting international standards and requirements (Keck & Sikking, 1999), rather than a means to generate evidence to tackle national problems. This is especially true under the umbrella of UN SDGs, clearly showing the growing influence of international non-governmental, non-state, actors and networks on measurement priorities and policies.

ORCID

Giulia Mugellini https://orcid.org/0000-0002-8213-4250
Jean-Patrick Villeneuve https://orcid.org/0000-0002-1164-7497
Marlen Heide https://orcid.org/0000-0003-4702-0896

ENDNOTES

[1] The UN Interagency and Expert Group on SDG Indicators (IAEG-SDGs) periodically classifies SDG indicators into three tiers on the basis of their methodological development and data availability across Member States (see Tier Classification for Global SDG Indicators https://unstats.un.org/sdgs/iaeg-sdgs/tier-classification/).

[2] https://unstats.un.org/sdgs/iaeg-sdgs/tier-classification/

[3] Report of the Inter-Agency and Expert Group on Sustainable Development Goals Indicators (E/CN.March 2, 2017*).

[4] Whether or not a measure reflects the underlying concept to be monitored (Drost, 2011).

[5] Other, more indirect methods for measuring corruption such as expert assessments and composite indices, entail a combination of subjective assessments, perceptions and other data that might be linked to corruption (UNODC-UNDP 2018, p. 20–25).

[6] Proportion of persons who had at least one contact with a public official and who paid a bribe to a public official or were asked for a bribe by those public officials, during the previous 12 months. Report of the Inter-Agency and Expert Group on Sustainable Development Goals Indicators (E/CN.March 2, 2017*).

[7] Proportion of businesses who had at least one contact with a public official and who paid a bribe to a public official or were asked for a bribe by those public officials, during the previous 12 months. Report of the Inter-Agency and Expert Group on Sustainable Development Goals Indicators (E/CN.March 2, 2017*).

[8] UNODC-UNDP (2018) project on "Producing methodological guidelines on the measurement of corruption at national level"

[9] United Nations bodies (and in some cases, other international organizations) responsible for compiling and verifying country data and metadata, and for submitting the data, along with regional and global

aggregates, to the United Nations Statistics Division (UNSD). https://www.sdg6monitoring.org/activities/roles-and-responsibilities/

[10] S2 and S3 in the Supporting Information present basic information on each of the identified survey projects.

[11] Metadata 16_05_02 https://unstats.un.org/sdgs/metadata/?text=&goal=16

[12] QUalityEvaluationSdgTarget16.5

[13] Metadata 16_05_02 https://unstats.un.org/sdgs/metadata/?text=&goal=16

[14] When measuring human behavior and social phenomena, it is frequent to hear also about construct validity as a dimension of accuracy. According to Drost (2011, p. 116) "construct validity refers to how well a concept, idea, or behavior – that is a construct – is translated or transformed into a functioning and operating reality, the operationalisation".

[15] In the majority of countries, this is considered as the legal adult age that allow individuals to interact with public officials for official administrative procedures.

[16] This ensures the reliability of surveys' results as it enables the probability of each sampling unit (individual) being included in the sample and, thus, the sampling weights used in the estimation phase to be calculated properly (UNODC-UNDP, 2018, p. 74).

[17] Face-to-face and CATI are those survey modes presenting the highest response rates for corruption surveys.

[18] https://www.access-info.org/blog/2020/07/17/open-sdg-data-key-agenda-2030/

[19] When the public official or private entity asks, solicits, or lets the citizen understand a bribe is necessary to obtain a specific service or avoid fees/fines.

[20] Proportion of persons who had at least one contact with a public official and who paid a bribe to a public official or were asked for a bribe by those public officials, during the previous 12 months. Report of the Inter-Agency and Expert Group on Sustainable Development Goals Indicators (E/CN.March 2, 2017*).

[21] Proportion of businesses who had at least one contact with a public official and who paid a bribe to a public official or were asked for a bribe by those public officials, during the previous 12 months. Report of the Inter-Agency and Expert Group on Sustainable Development Goals Indicators (E/CN.March 2, 2017*).

[22] As stated in the UNODC-UNDP Manual (2018, p. 83): "In a corruption survey among the general population, it is preferable to use a screening question based on public officials, as this ensures more comprehensive coverage of citizens' bribery experiences. That is because contact with public officials does not always take place in relation to a specific procedure (e.g., when dealing with police officers or teachers). Furthermore, citizens seem to recall contacts and dealings with people (in this case public officials) better than occurrences related to procedures".

[23] Information related to Suggestions are not available for all identified surveys. When the information is available, it appears that large generic surveys are more likely to stick with these suggestions, because they usually have big sample and they need to pilot the questionnaire and methodology or to perform cognitive testing. Besides this structural reason, there are also reasons linked to the available resources.

[24] I.e.; V&S_P_LUX; VS_P_KEN; VS_P_EGY; VS_P_TZA; VS_P_UGA; VS_P_RWA; SCC_P&B_CPV; VOCS_P_ZAF.

[25] "Thinking about these contacts in the past 12 months has anyone in (OUR COUNTRY) asked you or expected you to give a gift, favor, or extra money for his or her services". (European Commission, 2017, p. 120).

[26] "Has anyone asked or expected someone from your company to pay a bribe for any of the following permits or services?". (European Commission 2015, p. 159).

[27] "In some countries, there is a problem of corruption among government or public officials. During 2004, did any government official, for instance a customs officer, a police officer, a judge or inspector in your country ask you, or expected you to pay a bribe for his or her services".

REFERENCES

Allen, C., Metternicht, G., & Wiedmann, T. (2018). Initial progress in implementing the sustainable development goals (SDGs): A review of evidence from countries. *Sustainability Science*, *13*, 1453–1467. https://doi.org/10.1007/s11625-018-0572-3

Andersson, S. (2016). Beyond unidimensional measurement of corruption. *Public Integrity*, *19*(1), 58–76. https://doi.org/10.1080/10999922.2016.1200408

Batini, C., & Scannapieco, M. (2006). *Data Quality: Concepts*. Berlin Heidelberg: Springer-Verlag.

Bauhr, M., & Grimes, M. (2017). Transparency to curb corruption? Concepts, measures and empirical merit. *Crime Law Soc Change*, *68*, 431–458. https://doi.org/10.1007/s10611-017-9695-1

Beeri, I., & Navot, D. (2013). Local political corruption: Potential structural malfunctions at the central–local, local–local and intra-local levels. *Public Management Review*, *15*(5), 712–739. https://doi.org/10.1080/14719037.2012.707682

Blind P.K. (2011). Perspective of corruption metrics. Prepared for the workshop on Engaging Citizens to Counter Corruption for Better Public Service Delivery and Achievement of the Millenium Development Goals. 4th Session of the Conference of the State Partie (CoSP) to the United Nations Convention against Corruption (UNCAC), Morocco.

Bussell, J. (2015). Typologies of corruption: A pragmatic approach. In S. Rose-Ackerman & P. Lagunes (Eds.), *Greed, corruption, and the modern state* (pp. 21–45). Cheltenham: Edward Elgar.

Chabova, K. (2017). Measuring corruption in Europe: Public opinion surveys and composite indices. *Quality & Quantity*, *51*(4), 1877–1900. https://doi.org/10.1007/s11135-016-0385-3

Chengalur-Smith, I. N., Ballou, D. P., & Pazer, H. L. (1999). The impact of data quality information on decision making: An exploratory analysis. *IEEE Transactions on Knowledge and Data Engineering*, *11*(6), 853–864.

Cooley, A., & Snyder, J. (2015). *Ranking the world: Grading states as a tool of global governance*. Cambridge: Cambridge University Press.

De Graaf, G. (2007). Causes of corruption: Towards a contextual theory of corruption. *Public Administration Quarterly*, *31*(1–2), 39–86.

Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and Perspectives*, *38*(1), 105–123.

Dupuy, K., & Neset, S. (2018). The cognitive psychology of corruption. Micro-level explanations for unethical behavior. U4 Issue, *2018*, 2.

European Commission. (2015). Flash Eurobarometer 428 "Businesses' attitudes towards corruption in the EU" Report. *European Commission*, 2015.

European Commission. (2017). Special Eurobarometer 470 – Corruption – Report. *European Union*, *2017*. https://doi.org/10.2837/513267

Eurostat (2018). Sustainable development in the European Union. In: Monitoring Report on Progress Towards the SDGs in an EU Context. Publications Office of the European Union, Eurostat statistical books, Luxembourg. https://doi.org/10.2785/221211.

Eurostat. (2020). Sustainable development in the European Union overview of progress towards the SDGs in an EU context. *European Union*, *2020*. https://doi.org/10.2785/634624

Fredriksson, A. (2014). Bureaucracy intermediaries, corruption and red tape. *Journal of Development Economics*, *108*, 256–273. https://doi.org/10.2139/ssrn.3006070

Gould, D. J. (1991). Administrative corruption: Incidence, causes, and remedial strategies. In A. Farazmand (Ed.), *Handbook of comparative and development public administration* (pp. 467–480). New York: Dekker.

Graycar, A. (2015). Corruption: Classification and analysis. *Policy and Society*, *34*, 87–96.

Gupta, A. (1995). Blurred boundaries: The discourse of corruption, the culture of politics, and the imagined state. *American Ethnologist*, *22*(2), 375–402.

Hammer, C. L., Kostroch, D. C., Quiròs, G., & Ducharme, L. M. (2017). Big Data: Potential. *Challenges, and Statistical Implications. IMF Staff Discussion Note*, *SDN/17/06* (Vol. *2017*, pp. 6–30). https://www.elibrary.imf.org/view/journals/006/2017/006/article-A001-en.xml

Herrera, J., Razafindrakoto, M., & Roubaud, F. (2007). Governance, democracy and poverty reduction: Lessons drawn from household surveys in sub-Saharan Africa and Latin America. *International Statistical Review*, *75*(1), 70–95. https://doi.org/10.1111/j.1751-5823.2006.00005.x

Heywood, P. M. (2017). Rethinking corruption: Hocus-pocus, locus and focus. *The Slavonic and East European Review*, *95*(1), 21–48.

Heywood, P. M., & Rose, J. (2014). Close but no cigar: The measurement of corruption. *Journal of Public Policy*, *34*, 507–529.

Holmes, L. (2015, 2015). *Corruption: A very short introduction*. Oxford: Oxford University Press.

Huberts, L. W. J. C. (1998). What can be done against public corruption and fraud: Expert views on strategies to protect public integrity. *Crime, Law and Social Change*, *29*, 209–224.

Jancsics, D. (2019). Corruption as resource transfer: An interdisciplinary synthesis. *Public Administration Review*, *79*(4), 523–537. https://doi.org/10.1111/puar.13024

Jandl, M. (2017). Towards the monitoring of goal 16 of the United Nations' Sustainable Development Goals (SDGS) a study of the selection, rationale and validity of indicators with suggestions for further improvements. Helsinki 2017: European Institute for Crime Prevention and Control, affiliated with the United Nations (HEUNI).

Johnston, M. (2005). *Syndromes of corruption: Wealth, power, and democracy*. Cambridge: Cambridge University Press.

Johnston, M. (2013). *Corruption, contention and reform: The power of deep democratization*. Cambridge: Cambridge University Press.

Kaufmann, D., & Vicente, P. C. (2011). Legal corruption. *Economics and Politics*, *23*(2), 195–219. https://doi.org/10.1111/j.1468-0343.2010.00377.x

Keck, M. E., & Sikking, K. (1999). Transnational advocacy networks in international and regional politics. *International Social Science Journal*, *51*, 89–101.

Knack, S. (2007). Measuring corruption: A critique of indicators in Eastern Europe and Central Asia. *Journal of Public Policy*, *27*, 255–291.

Lafortune, G., Fuller, G., Moreno, J., Schmidt-Traub, G., & Kroll, C. (2018). SDG Index and Dashboards: Detailed Methodological Paper.

Lynch, J. P. (2006). Problems and promise of victimization surveys for cross-national research. *Crime and Justice*, *34*(1), 229–287.

MacFeely, S. (2019). The big (data) bang: Opportunities and challenges for compiling SDG indicators. *Global Policy*, *10*(S1), 121–133.

Maciel, G. G., & de Sousa, L. (2018). Legal corruption and dissatisfaction with democracy in the European Union. *Social Indicators Research*, *140*, 653–674. https://doi.org/10.1007/s11205-017-1779-x

Malito, D. (2014). Measuring corruption indicators and indices. *Robert Schuman Centre for Advanced Studies Research Paper*, *2014*(13), 1–26. https://doi.org/10.2139/ssrn.2393335

Marquette, H., & Pfeiffer, C. (2015). Corruption and collective action. The Developmental Leadership Program (DLP) and U4 Research Paper 32.

Merry, S. E. (2019). The sustainable development goals confront the infrastructure of measurement. *Global Policy*, *10*(S1), 146–148. https://doi.org/10.1111/1758-5899.12606

Miola, A., & Shiltz, F. (2019). Measuring sustainable development goals performance: How to monitor policy action in the 2030 Agenda implementation? Ecological Economics. https://doi.org/10.1016/j.ecolecon.2019.106373

Mugellini, G. (2018). Critical review of existing best practices to measure the experience of corruption. UNODC-INEGI Center of Excellence in Statistical Information, Government, Crime, Victimization and Justice, Working Paper Series http://www.cdeunodc.inegi.org.mx/unodc/wp-content/uploads/2020/02/Critical_Review_Corrupt_Measrmts.pdf.

Mugellini, G. (2020). Corruption. In P. Harris, A. Bitonti, C. Fleisher, & B. A. Skorkjær (Eds.), *The Palgrave encyclopedia of interest groups, lobbying and public affairs*. Cham: Palgrave Macmillan.

Mugellini, G., & Villeneuve, J. P. (2019). Monitoring the risk of corruption at international level: The case of the United Nations sustainable development goals. *European Journal of Risk Regulation*, *10*(1), 201–207. https://doi.org/10.1017/err.2019.16

Mungiu-Pippidi, A., & Dadašov, R. (2016). Measuring control of corruption by a new index of public integrity. *European Journal on Criminal Policy and Research*, *22*(3), 415–438. https://doi.org/10.1007/s10610-016-9324-z

Navot, D. (2014). The concept of political corruption. *Public Integrity*, *16*(4), 357–374.

OECD (2015). Consequences of Corruption at the Sector Level and Implications for Economic Growth and Development. Abingdon: OECD. https://doi.org/10.1787/9789264230781-en.

OECD (2017). Measuring Distance to the SDG Targets: An Assessment of Where OECD Countries Stand. Paris. http://www.oecd.org/sdd/OECD-Measuring-Distance-to-SDG-Targets.pdf.

OECD (2007). Bribery in Public Procurement. Methods, Actors and Counter-Measures. Abingdon: OECD.

OECD (2008). Corruption. A glossary of international and criminal standards. https://www.oecd.org/daf/anti-bribery/41194428.pdf

Rose, J. (2018). The meaning of corruption: Testing the coherence and adequacy of corruption definitions. *Public Integrity*, *20*, 220–233.

Rose-Ackerman, S. (1978). *Corruption: A study in political economy*. New York: Academic Press.

Rubio, D.F. & Andvig, E. (2019). Serious about sustainability? Get serious about corruption https://www.weforum.org/agenda/2019/09/serious-about-sustainability-get-serious-about-corruption/

Ryvkin, D., Serra, D., & Tremewan, J. (2017). I paid a bribe: An experiment on information sharing and extortionary corruption. *European Economic Review*, *94*, 1–22.

Sachs, J., Schmidt-Traub, G., Kroll, C., Lafortune, G., & Fuller, G. (2018). *SDG index and dashboards report 2018*. New York: Bertelsmann Stiftung and Sustainable Development Solutions Network (SDSN).

Sachs, J., Schmidt-Traub, G., Kroll, C., Lafortune, G., Fuller, G., & Woelm, F. (2020). *The sustainable development goals and COVID-19*. *Sustainable development report 2020*. Cambridge: Cambridge University Press.

Sequeira, S. (2012). Advances in measuring corruption in the field. In D. Serra & L. Wantchekon (Eds.), *New advances in experimental research on corruption* (pp. 145–175). Bingley: Research in Experimental Economics, Emerald Group Publishing Limited. https://doi.org/10.1108/S0193-2306(2012)0000015008

Thatcher, R. (2010). Validity and reliability of quantitative electroencephalography (qEEG). *Journal of Neurotherapy*, *14*, 122–152.

The World Bank. (2000). *Anticorruption in transition: A contribution to the policy debate*. Washington, DC: World Bank.

The World Bank (2003). Youth for Good Governance. Module III. Introduction to Corruption. http://info.worldbank.org/etools/docs/library/35970/mod03.pdf

Transparency International. (2013). *Global corruption report: Education*. London: Routledge.

UNESCO Institute for Statistics (UIS) (2017). The Quality Factor: Strengthening National Data to Monitor Sustainable Development Goal 4. SDG 4 DATA DIGEST 2017.

United Nations (2019). United Nations National Quality Assurance Frameworks Manual for Official Statistics. New York, 2019.

United Nations Office on Drugs and Crime (2006). Legislative guide for the implementation of the United Nations Convention against Corruption.

United Nations Office on Drugs and Crime (UNODC), United Nations Development Programme (UNDP) and UNODC-INEGI Center of Excellence in Statistical Information on Government, Crime, Victimization and Justice. (2018). *Manual on corruption surveys*. Vienna: UNODC.

United Nations Statistical Commission (UNStats) (2017). Guidelines and Best Practices on Data Flows and Global Data Reporting for Sustainable Development Goals. https://unstats.un.org/sdgs/files/meetings/iaeg-sdgs-meeting-06/20171108_Draft%20Guidelines%20and%20Best%20Practices%20for%20Global%20SDG%20Data%20Reporting.pdf

Villeneuve, J. P., Mugellini, G., & Heide, M. (2019). International anti-corruption initiatives: A classification of policy interventions. *European Journal on Criminal Policy and Research*, *26*–431–455. https://doi.org/10.1007/s10610-019-09410-w

Wickberg, S., & Mugellini, G. (2020). The evolving nature of evidence as used within the international anti-corruption community. In *Laboratory for interdisciplinary evaluation of public policies (LIEPP)*. Paris: Laboratory for Interdisciplinary Evaluation of Public Policies.

Wysmulek, I. (2019). Using public opinion surveys to evaluate corruption in Europe: Trends in the corruption items of 21 international survey projects, 1989–2017. *Quality & Quantity*, *53*(5), 2589–2610.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Mugellini, G., Villeneuve, J.-P., & Heide, M. (2021). Monitoring sustainable development goals and the quest for high-quality indicators: Learning from a practical evaluation of data on corruption. *Sustainable Development*, *29*(6), 1257–1275. https://doi.org/10.1002/sd.2223