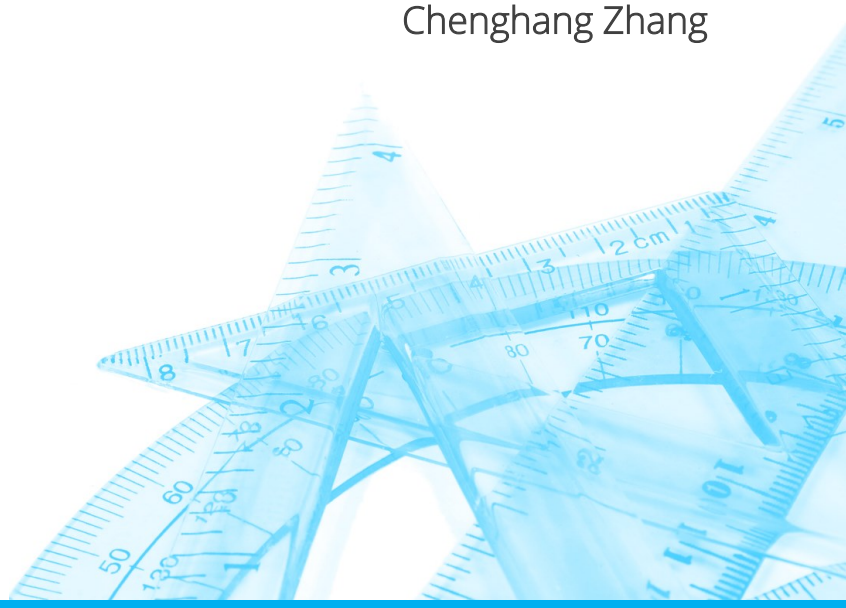
Assessing Sustainability: Frameworks and Indices

Sustainability Metrics White Paper Series: 3 of 3

March 2015

Dr. Dong Guo
Kelsie DeFrancia
Meiyi Chen
Brian Filiatraut
Chenghang Zhang



RESEARCH PROGRAM ON
Sustainability Policy and Management
EARTH INSTITUTE | COLUMBIA UNIVERSITY

The Earth Institute Research Program on Sustainability Policy and Management provides a rigorous analytic base to help inform sustainability decision-making. Our research addresses the fundamental issues facing professionals and policy makers implementing sustainability strategies. We seek to better understand the mechanisms behind sustainability management, in order to develop and promote more effective public policies and organizational practices. We analyze sustainability strategies and initiatives, examine methods of valuing sustainability practices, and study the impact of policies that stimulate sustainability innovations and trends. The goal of the program is to develop models to overcome barriers to institutionalizing sustainability in organizational operations. We aim to hasten the integration of sustainability principles in the management of organizations by providing the data necessary for decision-making.

spm.ei.columbia.edu

RESEARCH PROGRAM ON
Sustainability Policy and Management
EARTH INSTITUTE | COLUMBIA UNIVERSITY

Table of Contents

Assessing Sustainability: Frameworks and Indices

Executive Summary

The ability to accurately assess and track sustainability is crucial for achieving sustainable development goals at every level. There are many types of tools for sustainability assessment, at the core of which are indicators and indices. By presenting a number of common sustainability frameworks and outlining the underlying steps for constructing sustainability indices, this paper aims to contribute to the overview and discussion of sustainability measurement, which is often inconsistent and confusing. This paper also analyzes how sustainability frameworks can assist in the selection of indicators, which is often the most important yet most inconsistent step in constructing an index, and points out the need for future research to develop a dynamic and objective process of indicator selection for both frameworks and composite indices.

Introduction

Sustainability has become a widely acknowledged objective to incorporate into the daily operations of companies and organizations. The ability to accurately assess and monitor quantitative measures of sustainability is crucial for achieving sustainable development goals at every level. There are generally three types of tools that have been adopted by researchers and organizations to measure sustainability: 1) product-related assessment tools that focus on material and/or energy flow of a product or service; 2) integrated assessment tools with the aim for policy or project implementation; and 3) indicators and indices (Ness, Urbel-Piirsalu, Anderberg, & Olsson, 2007). The product-related assessment delineates flows of energy or materials concerning production and consumption of goods and services and aims to identify risks and inefficiencies in the production processes. Well-known product instruments include Life Cycle Assessment (LCA) and the product Material Intensity Analysis developed by the Wuppertal Institute for Climate, Environment and Energy. Integrated assessments focus on evaluating complex scenarios using tools such as conceptual modeling, multi-criteria analysis, risk and uncertainty analysis, and cost-benefit assessment, and are often adopted to aid policymaking. Indicators are simple measures that describe current states of a company or a region with respect to some sustainability category, and indices are the results of those indicators aggregated into a single measure. Indicators and indices are at the foundation of sustainability assessment and are the focus of this paper.

Sustainability indicators are able to transform a vast amount of information about our complex environment into concise, policy-applicable and manageable information. There is a very large universe of indicators to measure the sustainability performance of an entity, but the critical question is what to use and how many indicators should be evaluated. Sustainability indicators are either presented in a structured framework that can be used to isolate and report on relevant indicators, or aggregated towards a composite index or score/rating. This paper discusses sustainability assessment using both non-integrated indicators (presented in a framework) and integrated indicators or indices (the consolidation of two or more indicators), given that the academic literature has documented frequently the necessity to quantify concepts of sustainability into metrics or indices.[1]

The adaptability and evolution of frameworks and indices are essential because our collective understanding of sustainability changes rapidly. This evolution involves a change in indicators, entities

[1] See (Tanzil & Beloff, 2006), (Szekely & Knirsch, 2005), and (Azapagic & Perdan, 2000) as examples.

and interpretations of results, but research on the proper frequency of framework revision hasn't yet been conducted.[2] This whitepaper therefore aims to summarize the most commonly used frameworks, outline the methods frequently used in constructing indices, such as standardization, weighting, and aggregation, and analyze how sustainability frameworks can assist in the selection of indicators, which is often the most important yet most inconsistent step in constructing an index.

Sustainability Frameworks

There are myriad frameworks developed and proposed by various organizations and researchers, which are qualitative presentations of sustainability used to isolate and report on relevant indicators. The most influential framework, upon which many other sustainability frameworks have been built, is the Triple Bottom Line, which was proposed by John Elkington (1998) to represent social, environmental, and economic pillars of sustainability. He argued that corporations should perform above three bottom lines: profit and loss account, people account, and planet account. Measuring sustainability based on these three segments does not indicate that organizations must maximize returns on these dimensions, but rather that the social, environmental and economic performances of an entity should be integrated and analyzed as a whole (Moneva, Archel, & Correa, 2006). The Triple Bottom Line currently represents one of the most widely accepted framework foundations to evaluate an institution's performance in sustainable development.

Perhaps the most common example of a Triple Bottom Line framework is the Global Reporting Initiative (GRI), which was developed by the Coalition for Environmentally Responsible Economies (CERES) with the United Nations Environment Programme (UNEP) in 2000. The third version of the GRI guideline considers a set of 84 indicators across three pillars—social, environmental and economic—with the largest emphasis placed on the social and environmental aspects (Das & Das, 2014). In addition, the United Nations Commission for Sustainable Development devised a Theme Indicator Framework. By incorporating institutional aspects in addition to the social, environmental, and economic components, this framework aims to evaluate the process of sustainable development from the government's perspective. The four components comprise indicators measuring 38 sub-themes, such as equity and health for the social sector, land and ocean for the environment sector, economic structure for the economic sector, and institutional capacity for the institutional aspect.

Also related to the Triple Bottom Line foundation, another common framework for sustainable development indicators includes variations of the causal chain or "stress response model". The pressure-state-response (PSR) framework lays the groundwork for many subsequent sustainability measurement models by describing the process in which human activities create particular pressures and thus cause responses from societal, economic, environmental or social agents (Singh, Murty, Gupta, & Dikshit, 2009). The PSR model delineates a procedure of how people influence their surrounding environment and how it in turn reacts. This model can then be useful to organizations for measuring the impact of their actions on the environment, and for evaluating the impacts of different management response options.[3]

[2] See (Searcy, 2012).

[3] The PSR model was later revised to a more comprehensive version, which is the driver-pressure-state-impact-response (DPSIR) model. It has the same methodological foundation with the original PSR model.

Aside from considering all aspects of sustainability, frameworks have also been developed to provide qualitative evaluation of composite environmental, social and economic indicators. For example, the Barometer of Sustainability (Prescott-Allen, 1995) was developed to evaluate the environmental and social aspects of sustainability at the same time; the Ecological Footprint (Wackernagel & Rees, 1996) is specified by an area of land needed to produce enough food, water, energy, as well as to dispose waste, for a person, a product or a city; and the Eco-Efficiency Framework (World Business Council for Sustainable Development [WBSCD], 1999) helps businesses assess their sustainable development using combined economic and environmental indicators that are related to the business world. Lastly, there are even frameworks that solely focus on one aspect of sustainability, such as the indicator framework developed by the Lowell Center for Sustainable Production, which aims to present a system of environmental sustainability indicators specifically designed for the production process. The indicator system contains five levels, which are facility compliance/conformance indicators, facility material use and performance indicators, facility effect indicators, supply chain and product life-cycle indicators, and sustainable system indicators.

Finally, there are various other frameworks that stand independent from these common frameworks for sustainability, such as the Competing Values Framework, and the Approach, Deployment, Results, and Improvement (ADRI) assessment matrix.[4] Overall, sustainability frameworks aim to measure sustainability primarily by providing qualitative evaluations of processes or selected composite environmental, social and economic indicators. This contrasts with sustainability indices that aim to provide quantitative evaluation of sustainability through composite indicators.

Sustainability Indices

Indicators are quantitative measures that represent a state of economic, social, and/or environmental development in a defined region (Ness et al., 2007). When indicators are aggregated, the resulting composite measure becomes an index. Therefore, indices or composite indicators are integrated indicators in the sense that more than one dimension of sustainability are aggregated. For example, the Index of Sustainable Economic Welfare (ISEW) and the Environmental Sustainability Index (ESI) move beyond economic accounting and encompass numerous nature-society dimensions. ESI in particular aggregates 68 indicators from five categories of mostly environmental sustainability aspects, but also from social and institutional dimensions.

Even though the population of sustainability indices has grown significantly over the past few decades, there is a common system of converting a set of indicators to complex indices. First, a group of indicators are selected depending on the purpose of the index. Scores are assigned to each indicator and those scores are standardized and assigned relative weightings against each other. The final step of compiling a complex index is to aggregate the results of all indicators using various mathematical calculations. The specific methods used in each step are outlined in the following sections.

Indicator Selection

Most indices include indicators based on what they seek to measure and report to their target audience. The chosen indicators of each index will cover the basic, essential, and comprehensive aspects of targets' sustainability, which would vary with the nature of the target audience. Multiple methods can be used to categorize indicators for indices to measure various aspects of sustainability.

[4] See (Robbins & Page 2012), (Pojasek, 2007), and (Newman, 2007) for details.

In addition, it is rare for an index to utilize a universal rating method across all sectors/ industries. Changes between industries involve variation in the indicators used, categories scored, or both of these features. When the reasoning for the variation between indicators used for different industries is provided, indices typically maintain that not all indicators are material to all industries and, therefore, inclusion of the same indicators across all industries would not produce relevant information. For example, carbon emissions from fuel combustion is a more critical indicator for the shipping industry than the software industry.

The nature of the target audience is not the only rationale of indicator selection, and there is a range of rules suggested by the literature on how an indicator should be selected for inclusion in the construction of an index. For example, the Environmental Performance Index developed by the Yale Center for Environmental Law and Policy and the Center for International Earth Science Information Network at Columbia University suggests: (1) less may be more; (2) establish causal connections between indicators and index goals; (3) if a perfect indicator doesn't exist, look for a proxy (especially for some complex factors); (4) consider relevant spatial and temporal scales; (5) keep in mind the target audience; (6) establishing a set of principles or criteria can help select indicators; (7) conduct a cost-benefit analysis to review resource and time constraints for data collection; and (8) quality of data counts (Hsu et al., 2013).

However, the selection processes are still arbitrary without a set of standardized and widely adopted requirements. Most indices relied upon in-house "expert" panels/employees to determine key aspects and key indicators, especially for those measuring corporate sustainability performance. For example, there are no clearly articulated methodologies on indicator selections for indices such as the Physical Quality of Life Index, Sustainable Process Index, and Living Planet Index. However, this problem could be mitigated through the use of objective external information for data confirmation, indicator guidelines, general oversight, result comparisons, and process auditing.

Standardization

Score standardization—also called transformation or normalization in some cases—is to transform the original value of all indicators into a more generic value for the convenience of aggregation, and is commonly used by indices. The most widely used method is high performance benchmarking, which follows the concept of the equation:

$$Standardized\ Score = \frac{Actual\ Value - Minimum\ Value}{Maximum\ Value - Minimum\ Value}$$

The percentage results could then be converted into 0-10, 1-10, 0-100, or other ranges on a case-by-case basis. The minimum value can be the worst observed value of: 1) the target in a period of time (this means it is based on the target's own minimum demonstrated performance), or 2) all potential targets (this emphasizes comparisons with other targets). Maximum value could be calculated similarly.

In some cases, before standardization, indicator values may need to be pre-transformed. If an indicator contains scale effects, it needs to be standardized into a comparable form using generic denominators, such as population or GDP (Hsu, Johnson, & Lloyd, 2013). For instance, the Human Development Index (HDI) is population adjusted. In addition, some economic indicators, such as the Gross National Income, needs to be converted to constant terms, typically by adjusting for the rate of inflation.

Weighting and Aggregation

Similar to the selection of indicators, most weights are assigned arbitrarily, determined by in-house panels or employees. However, there are cases where indicators in some indices are weighted based on initial statistical calculation, such as Principal Components Analysis (PCA) and Factor Analysis (FA), to reflect indicators' relative contribution toward explaining the variance associated with changes over time. Principal Component Analysis converts a set of possibly correlated variables of observations into a set of linearly uncorrelated variables (principal components). Factor Analysis searches for unobserved variables (factors) that are reflected by observed variables in a greater amount. Both statistical methods could help to identify key indicators with their relative importance (or contribution). But PCA and FA, for weighting purposes, is only useful when it is to correct for overlapping information carried by correlated indicators; they cannot show the theoretical or policy importance of those indicators. Thus, when computing several indicators into a composite sub-dimension score, PCA and FA could be used to give weights.

In addition, Standard & Poor proposed several weighting methodologies in terms of the capitals of corporations. One of those, for example, which they use for their ESG (Environmental, Social, and Governance) India Indices, is called the Score Weight Factor, which is a weight that considers the stock's initial ESG score, stock share, stock price and their investable weight factor.

Un-weighted aggregation indicates either no weighting or evenly weighted aggregation. This simple averaging across indicators is sometimes used because it can be understood easily and more complex models do not appear to offer any advantages to the expression or utility of the index (Pratt, 2004). Also, this methodology avoids the dilemma of having to rigorously justify any evaluations of the relative importance of the indicators.

Once weight is given to each indicator (or in some cases, none), mathematical operation is performed to combine the indicators into a single value. Major aggregation methods adopted by indices include the basic mathematical functions of addition, subtraction, multiplication, and division. Arithmetic (taking averages across all indicator values) and geometric (using the product of indicator values instead of the sum) means appear to be the most commonly used methods when it comes to aggregating many indicators.

Another way to aggregate indicators is by the "veto method"—a lower score in one category overriding a higher score for another category. This method is used for constructing the Ecosystem Wellbeing Index (EWI), together with the average method, when aggregating the components and indicators. One major drawback for the veto method is that it reduces transparency by hiding the actual contributions of the overridden components or indicators in the aggregated value.

Discussion

The focus of this paper is on indicators and indices, the first tier in sustainability assessment. By presenting a number of common sustainability frameworks and outlining the underlying steps for constructing sustainability indices, this paper aims to contribute to the overview and discussion of sustainability measurement, which is often inconsistent and confusing. The inconsistency originates from the arbitrary selection of indicators, and the subsequent weighting criterion. Frameworks can effectively serve as guidelines for selecting indicators either for disaggregated 'dashboards' or for composite indices. They can provide a standard process for identifying the most critical measures of success when choosing indicators.

Concerns with Indicator Selection

The primary concern with index construction is the evidence of reliance on compartmentalized expert opinion for the indicator selection process and weighting system of sustainability performance indices. Although outside sources are utilized as resources for some cases, each measuring model could not be replicated without the compliance of the initial creators or managers. Not only does this limit the transparency of these indices, but it also limits their usefulness in direct comparison to other metric systems and convincingly showing the real sustainability performance of the entity. The reliance on expert opinion is also likely tied to the minimal usage of advanced and more objective aggregation techniques when compiling scores. The private companies creating these indices may be more interested in maintaining proprietary methods of analysis rather than utilizing universally accepted methodologies, as well as some cases for governmental or societal use.

Most indices develop industry-specific scores. This approach likely best accounts for the diversity of issues covered within "sustainability" and the manner in which these issues are critical, or less important, to a specific sector. The sector specificity of these ratings may be more beneficial than a highly generalized and universally applicable rating system, but cross-sector comparisons are impossible using this approach. Also, the use of surveys completed by the company being evaluated provides a questionable data source, as incentives exist for misrepresentation of data or the possibility of incomplete provision of data.

The reliance on proprietary expert opinion of all studied ratings systems implies that the universe of private sector ratings systems is highly compartmentalized and lacks an open discussion of methodological information. This may be one of the impediments to a lack of consensus regarding what constitutes "sustainability" within the private sector and how best to measure it. Given that this dearth of information also makes detailed comparisons amongst rating systems impossible through direct analysis, the best current method of generating a universal sustainability rating for a company could be creating an aggregation system for the scores from all of the rating systems in which a company is listed. This would account for potential disparities amongst ratings systems, though would still not solve the issue of how to properly weight each rating. Comparison of an individual rating system's rankings to external metrics related to sustainability may also provide some universal measurement of validity for any one rating system.

Frameworks Guiding Indicator Selection

Frameworks can serve as a standard process for identifying the most critical issues when choosing indicators, and can therefore serve as guidelines for selecting indicators to quantitatively measure sustainability.

Prior to the selection process, it is critical to first define the overall purpose of the framework, as well as the system boundaries. These boundaries can be temporal, spatial or lifecycle boundaries (Lundin & Morrison, 2002). Temporal boundaries are of particular significance due to the fact that we can only measure progress towards sustainability, rather than describing its absolute state (Azapagic & Perdan, 2000). With the first two steps in place, sustainability frameworks can guide the identification and choice of environmental, social and economic indicators. Instead of actually providing quantitative benchmarks on various environmental or social aspects, frameworks equip policymakers with an overarching guiding principle when it comes to assessing human impacts on the environment and our responses.

In the frameworks that identified a clear process for the selection of indicators, each step was developed based on the guiding goal of the parent organization. No methodology was specified in the

development of any framework reviewed, aside from addressing the particular intended audience. However, common guiding principles adapted from the United Nations' Commission on Sustainable Development (UNCSD), the Wuppertal Institute, the Lowell Center for Sustainable Production, and the Wellbeing Assessment frameworks were consulted in drafting the following recommendations for selecting inclusive and meaningful indicators:

1) Form a diverse stakeholder working group;
2) Determine the purpose of the assessment;
3) Define a timeline for indicator selection, measurement, and reporting;
4) Be broad in covering all aspects of sustainable development;
5) Choose and make public the agreed-upon criteria for performance evaluation and indicator selection;
6) Commit to an open, participatory process of continuous evaluation and improvement focused on long-term performance; and
7) Review results and assess the implications and the process as a whole.

A common framework, such as the ones mentioned before, should aim to balance the various aspects of sustainability by covering different dimensions with relatively equal weights. However, it is necessary to point out that existing sustainability frameworks may not be generalized to be applicable to all industries or institutions. Krajnc and Glavič (2005) argue that while it is important to measure sustainability with multiple indicators, it can be difficult for decision-making and cross-comparison. Dahl (2012) claims that "while managers and policy makers may find a wide set of indicators useful, decision-makers and the general public prefer a limited set of 10–15 indicators of the most relevant trends." Therefore, not all indicators selected based on one framework will be appropriate for all organizations or kinds of analyses. Furthermore, integrated indicators – the consolidation of two or more indicators – are often not incorporated as part of sustainability frameworks.

Frameworks can serve as guidelines for selecting indicators to quantitatively measure sustainability.

Conclusion

Sustainability indicators are either presented in a structured framework or aggregated towards a composite index. The logic of a framework can often serve as the basis for aggregation and constructing composite indicators. However, frameworks vary drastically in their types or functions due to the diverse concept of sustainability itself. In general, sustainability frameworks provide qualitative presentation and grouping of large number of indicators, and can be more revealing and accurate than aggregated indices, while indices tend to be easy to use and more understandable by the general public. According to Singh et al. (2009), key dimensions of measurement for indicators and indices should consider factors such as: specific aspects of sustainability measured; methodologies used to construct the index; comparison of sustainability measures across space, over time, and in absolute or relative terms; perspectives from inputs or outputs; coherence and clarity; data availability over time and space; and flexibility for allowing changes. Integrated indicators and indices are more intricate than non-integrated measurements in many of these aspects, as they incorporate additional factors from the environmental or social perspectives.

...a dynamic and objective process of indicator selection is urgently needed...

Frameworks cannot adequately provide accurate measurement of sustainability as frameworks lack standardized and objective evaluations using numbers. As such, they often do not involve quantitative aggregation of data, in contrast to indices. Despite a balanced approach, problems with frameworks, such as validity, reliability, comparability, simplicity, and data availability, still have to be overcome (Singh et al., 2009).

Indices, although aggregated into a single number, do in fact integrate multiple dimensions. In addition, the construction of indices often involves standardization and aggregation of indicators that rely on scientific statistical methods. However, significant differences in index scores can be generated through different aggregation methods, even when applying these methods to the same dataset. Varying elements of sustainability measurement, such as assumptions, biases, and methodological disparities, can lead to different results for the same indices using the same set of data.

Furthermore, as often discussed, indices or rating systems can still be subjective, largely because there is no generally accepted procedure regarding normalization and weighting, and the choice of variables and weighting of indicators is arbitrary. If indices are weighted, those determinations can be made based on the opinions of sustainability experts, the priorities of the developer, or the opinions of people living in the system.

Although frameworks can provide some guidance for indicator selection, and aim to balance the various categories of sustainability by assigning relatively equal weights to all categories, the rankings generated by such a heuristic approach are generally far from robust. Therefore, a dynamic and objective process of indicator selection is urgently needed for both frameworks and composite indices, which should be the focus for future research on sustainability assessment.

References

Anand, S., & Sen, A. (1994). Human development index: methodology and measurement (No. HDOCPA-1994-02). Human Development Report Office (HDRO), United Nations Development Programme (UNDP).

Azapagic, A., & Perdan, S. (2000). Indicators of sustainable development for industry: a general framework. Process Safety and Environmental Protection, 78(4), 243-261.

Böhringer, C., & Jochem, P. E. (2007). Measuring the immeasurable—A survey of sustainability indices. Ecological Economics, 63(1), 1-8.

Dahl, A. L. (2012). Achievements and gaps in indicators for sustainability. Ecological Indicators, 17, 14-19.

Das, D., & Das, N. (2014). Sustainability reporting framework: Comparative analysis of global reporting initiatives and dow jones sustainability index. International Journal of Science, Environment and Technology, 3(1), 55-66.

Elkington, J. (1998). Cannibals with forks: The triple bottom line of 21st century business. Gabriola Island, BC: New Society Publishers.

Hsu, A., Johnson, L. A., & Lloyd, A. (2013). Measuring progress: A practical guide from the developers of the environmental performance index (EPI). New Haven: Yale Center for Environmental Law & Policy.

Krajnc, D., & Glavič, P. (2005). A model for integrated assessment of sustainable development. Resources, Conservation and Recycling, 43(2), 189-208.

Lundin, M., & Morrison, G. M. (2002). A life cycle assessment based procedure for development of environmental sustainability indicators for urban water systems. Urban water, 4(2), 145-152.

Moneva, J. M., Archel, P., & Correa, C. (2006). GRI and the camouflaging of corporate unsustainability. Accounting Forum, 30(2), 121-137.

Newman, J. (2007). An Organisational Change Management Framework for Sustainability. Greener Management International, 57, 65-75.

Ness, B., Urbel-Piirsalu, E., Anderberg, S., & Olsson, L. (2007). Categorising tools for sustainability assessment. Ecological Economics, 60(3), 498-508.

Pojasek, R. B. (2007). A framework for business sustainability. Environmental Quality Management, 17 (2), 81-88.

Pratt, C. (2004). The Environmental Vulnerability Index (EVI): Update & Final Steps to Completion (Vol. 369). Suba, Fiji Islands: South Pacific Applied Geoscience Commission.

Prescott-Allen, R. (1995). Barometer of sustainability: a method of assessing progress towards sustainable societies. Victoria, Canada: PADATA.

Robbins, G. E., & Page, M. (2012). Sustainability initiatives: a competing values framework. Competition Forum, 10(1), 176-181.

Searcy, C. (2012). Corporate sustainability performance measurement systems: A review and research agenda. Journal of Business Ethics, 107(3), 239-253.

Singh, R. K., Murty, H. R., Gupta, S. K., & Dikshit, A. K. (2009). An overview of sustainability assessment methodologies. Ecological Indicators, 9(2), 189-212.

Szekely, F., & Knirsch, M. (2005). Responsible leadership and Corporate Social responsibility: Metrics for Sustainable performance. European Management Journal, 23(6), 628-647.

Tanzil, D., & Beloff, B. (2006). Assessing impacts: Overview on sustainability indicators and metrics. Environmental Quality Management, 15(4), 41-56.

Wackernagel, M., & Rees, W. (1996). Our ecological footprint: Reducing human impact on the earth. Gabriola Island, BC: New Society Publishers.

World Business Council for Sustainable Development (WBCSD). (1999). Eco-efficiency indicators and reporting: Report on the status of the project's work in progress and guidelines for pilot application. Geneva, Switzerland: Authors.