IEc

Compliance Indexing Project

March 12, 2010

prepared for:

U.S. EPA, Office of Enforcement and Compliance Assistance

prepared by:

Industrial Economics, Incorporated

2067 Massachusetts Avenue

Cambridge, MA 02140

617/354-0074

INDUSTRIAL ECONOMICS, INCORPORATED

IEc

DISCLAIMER: This document was prepared under contract to the Environmental Protection Agency (EPA): Innovations Contract - Contract No. EP-W-10-002, WA 0-3. The views in this document are solely those of Industrial Economics and do not necessarily represent the opinions of the EPA or any of its employees.

TABLE OF CONTENTS

CHAPTER 1 | THEORETICAL DISCUSSION OF INDEXING

Introduction 1

Rates vs. Indexes 2

Case Study Indexes 4

CHAPTER 2 | REVIEW OF INDEX DEVELOPMENT APPROACHES

Introduction 10

Indicator Selection 10

Data Collection 13

Reporting 16

Summary of Index Construction Approaches 22

CHAPTER 3 | FINDINGS AND CONSIDERATIONS

Introduction 24

Use of Index 24

Selection of Indicators 25

Data Collection 26

Weighting of Indicators 26

REFERENCES 31

APPENDIX A: COMPLIANCE INDEXING CASE STUDY FACT SHEETS

CHAPTER 1 | THEORETICAL DISCUSSION OF INDEXING

INTRODUCTION

Compliance measurement is a crucial component of any regulator's toolbox. Understanding the performance of facilities in any given sector helps regulators adjust resources, efforts, and materials to achieve greater compliance. It also helps them identify facilities (or groups of facilities) that are poor performers and may need more enforcement or compliance assistance attention. To the extent that compliance indexes can lead to better dissemination of measurement information, their use can promote the adoption of more effective and cost-effective practices for environmental protection.[1]

Compliance rates can provide valuable information about the overall performance of a sector and can be more useful than outcome measures such as counts of inspections or enforcement actions. However, there are several definitional and usage issues with compliance rates.[2] For example, a high compliance rate for an industry can be misleading if the largest polluters remain out of compliance.[3] Compliance rates are not flexible instruments; they do not provide subtle measures of performance. In comparison, compliance indexing approaches can capture changes in performance more nuanced than a switch from compliance to noncompliance (or vice versa), and therefore may be better suited to distinguish between facilities' relative performance.

The primary purpose of this paper is to examine the current state of the science and use of environmental and other relevant compliance indexing approaches by federal, state, and local governments, associations, and international environmental agencies and organizations. In conducting our research, we relied primarily on a broad literature search for articles, reports, and other materials relevant to compliance indexing. We also conducted interviews with representatives of federal, state, and international agencies who have developed and applied compliance indexing approaches.

---

[1] Metzenbaum, Shelley. "Measure to Comply, Measure to Perform: A Government Performance White Paper". Cognos, Inc. February, 2006.

[2] "State Environmental Agency Contributions to Enforcement and Compliance". The Environmental Council of the States, Report to Congress, April 2001.

[3] "Recommendations on Performance Measurement for Environmental Enforcement and Authorities of Eastern Europe, Caucasus, and Central Asia." Eighth Annual Meeting of the EECCA Regulatory Environmental Programme Implementation Network. EAP Task Force, June 12/13, 2006.

RATES VS. INDEXES

A compliance rate is a binary measure of compliance; a facility is either in our out of compliance. A facility can only be marked as "in compliance" if it meets *all* of its regulatory requirements, or all of the indicators being measured. For example, the IRS measures its filing compliance rate with a compliance rate approach. This rate measures the number of tax returns timely filed as a percentage of the returns that were supposed to be filed (see equation below); in other words, how many taxpayers were in "compliance" with the IRS requirement to file returns on time.

$$Filing\ Rate = \frac{\text{Number of required returns filed timely}}{\text{Total number of returns required to be filed}}\ X\ 100$$

A compliance index, on the other hand, is a scoring system, expressed in the form of a number (e.g., 7.5 out of 10), or a percentage (e.g., 75 percent), representing the extent to which a given facility (or group of facilities) is (or are) complying with a pre-identified subset of the facility's (facilities') overall set of compliance obligations. Indexes generally involve aggregating information and sometimes include scoring mechanisms to provide a measure of the *extent* of compliance. For example, in a study to assess community compliance with water fluoridation standards, the authors created an index (called the community fluoridation compliance index – CFCI) to measure each water system's level of compliance. For a given system, all samples are compared to the optimal fluoride concentration, yielding a ratio of sample concentrations to optimal concentrations (see equation below). These values are summed and divided by the number of samples to provide a measure of the extent of compliance for each water system over a given time period.[4]

$$CFCI = \frac{\sum sample\ fluoridation\ concentration / optimal\ fluoridation\ concentration}{\#\ of\ samples}$$

Throughout this paper, we will discuss compliance measurement approaches that go beyond a simple compliance rate. For the purposes of this paper, "compliance indexing" will refer to any compliance measurement approach that is not a simple all or nothing measurement.

POTENTIAL BENEFITS OF INDEXING APPROACHES

Measuring compliance with an index has several potential benefits for regulators. In its examination of U.S. environmental indicators, the Government Accountability Office identified three benefits of using what it calls indicator sets (grouping of indicators): ability to assess subtle changes in environmental conditions and trends, communication of

[4] Kuthy, Raymond A. et al. "Use of a Compliance Index for Community Fluoridation." *Public Health Reports*, Vol. 102, No. 4, pp 415-420. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1477864/pdf/pubhealthrep00176-0069.pdf

complex environmental issues, and development of informed strategic plans for environmental management.[5]

The primary benefit of using a compliance index instead of a compliance rate is that an index allows regulators to capture a greater degree of subtlety in their compliance measurement efforts. Compliance indexes can measure compliance on a scale, not just as a binary measure. For example, if one facility is achieving one percent of its requirements, and another is achieving 99 percent of its requirements, these two facilities would both be marked as "noncompliant" under a compliance rate approach. Using an index, regulators can see that the performance of these two facilities is quite different.

Another benefit of using an index approach is that it allows regulators to compile certain indicators to compare performance across time or across facilities. By creating a "standardized" index, with common measures of compliance, regulators can remove the inconsistencies that may be measured in different years or across different groups. For example, the Common Measures Project was created in 2006 and funded through an EPA State Innovation Grant as a multi-state effort to evaluate the performance of targeted business sectors using common measures and statistical approaches. States developed regulatory and beyond compliance indicators and used statistical methods to estimate performance levels among small quantity generators (SQGs) of hazardous waste. By developing a subset of the most important indicators to measure and aggregating scores across facilities, the participating states were able to assess *gradations* in performance for the highest priority issues.[6]

Similarly, if a facility has made changes to move towards compliance, but has still not reached *full* compliance, a compliance rate would not capture that progress. Using an index approach gives regulators a way to track progress of facilities on a shorter time frame because it measures the extent of compliance in a more incremental way. Colorado's Department of Public Health and Environment's experience in measuring compliance for large quantity hazardous waste generators (LQGs) is an excellent example of how moving beyond a simple compliance rate can lead to better detection of progress. In 1999, the compliance rate among LQGs was around 70 percent. After a multi-year effort to increase inspections with the goal of improving compliance, the agency found compliance rates had not changed, leading regulators to initially believe that the inspection effort failed to improve compliance. Upon further investigation, however, the state found that, while the same number of facilities had at least one violation and continued to be counted as noncompliant, the overall number of violations had decreased – in 2000, noncompliant facilities had eight or more violations and in 2004 the average was four violations per facility. Thus, using a more subtle measure of compliance, regulators were able to see progress in the sector.[7]

[5] "Environmental Indicators: Better Coordination is Needed to Develop Environmental Sets that Inform Decisions". Government Accountability Office. GAO-05-52, November, 2004.

[6] Steven DeGabriele, Susan Peck and Tara Acker. The States Common Measures Project. June 19, 2009.

[7] Shewmake, Tiffin. "Using Compliance Rates to Manage." ECOStates Journal, Fall 2004. pp. 17-22.

We note that compliance indexes may not always be the best approach for measuring compliance. For instance, if regulators need to know only whether facilities are in full compliance or not, a compliance rate achieves that goal. Since creating compliance indexes can require significant resources, it only makes sense to pursue this approach where compliance rates are not likely to sufficiently meet an agency's needs. Furthermore, if the data used to compile an index do not provide substantial additional information about facility performance, little will be gained by using a compliance indexing approach. For example, in a study comparing the effectiveness of monitoring and enforcement actions, the authors found that noncompliance indexes provided little information beyond the information contained in basic noncompliance rates because in their dataset very few facilities had multiple repeated instances of noncompliance. Therefore, the noncompliance indexes essentially reproduced the basic noncompliance rates.[8]

Finally, it can be more difficult to communicate the results of a compliance index than a compliance rate, as the construction of an index is inherently more complicated. Compliance rates are readily understood: a facility is in compliance or it is not. A compliance index, when reported, is a numerical value that must be explained. Regulators must convey which indicators are included, how the index was constructed (e.g., weighting schemes), and what the results mean (i.e., what represents a bad score or a good score). While an index may provide regulators with more information about facility or sector performance, it is important to recognize the tradeoff an index approach requires with respect to the communication of results.

CASE STUDY INDEXES

Through our research, we identified 13 case study examples of indexes we use to illustrate some of the concepts discussed in this paper. Nine of these case studies focus on environmental issues. The remaining four cases are included because they provide relevant examples of indexing approaches. These examples are evenly split between those that directly measure compliance with regulatory requirements and those that capture more general performance measured against a goal. Three of the case study indexes measure both compliance with regulatory requirements and performance of non-regulatory, beyond compliance activities. Some of the indexes were developed by regulatory agencies, and others were constructed as part of an academic effort. This list of case studies is by no means exhaustive. Exhibit 1-1 below provides a summary table of these 13 case study indexes.

[8] Shimshack, Jay P. "Monitoring, Enforcement, & Environmental Compliance: Understanding Specific & General Deterrence: Comparative Analysis of Monitoring and Enforcement Impact Measures". Prepared for the U.S. EPA Office of Enforcement and Compliance. June 2009.

EXHIBIT 1-1: SUMMARY OF CASE STUDIES

INDEX NAME	AGENCY/ ORGANIZATION	WHAT IS MEASURED	INDEX DESCRIPTION	INTERVIEW CONTACT AND TITLE (IF APPLICABLE)	CITATION OR WEBSITE
Academic Performance Index	California School System	Academic performance	The performance index is a calculation that measures achievement and test performance at the 3rd, 4th, 5th, 6th, 7th, 8th, and 10th grade levels based upon the number of students at each performance level. The index is calculated by assigning a weighted score to each performance level. The percentage of students at each performance level is then multiplied by their respective weight, and the totals for each performance level are summed to get the building's overall performance index score.		http://www.cde.ca.gov/ta/ac/ap/ California Department of Education. "2008-09 Academic Performance Index Report: Information Guide." May 2009. LRC 2006-07 Documentation of Performance Index.
Common Measures Project	Massachusetts Department of Environmental Protection, Northeast Waste Management Officials Association, and U.S. EPA	Performance levels of Small Quantity Generators (SQGs) on a series of compliance and beyond compliance indicators	This project is designed to support state efforts to develop and use common measures of environmental performance for one or more business sectors/groups. It was funded by the 2005-2006 EPA State Innovations Grant Program (SIG). Ten states participated in this effort, which involved calculating two types of scores: • SQG mean facility score – state-wide average score for facilities achieving a particular set of indicators • Achievement rate – represents the state-wide percentage of facilities that achieved a specific indicator	Susan Zampaglione, CT Department of Environmental Protection	http://www.newmoa.org/hazardouswaste/measures/index.cfm Steven DeGabriele, Susan Peck and Tara Acker. The States Common Measures Project. June 19, 2009.

INDEX NAME	AGENCY/ ORGANIZATION	WHAT IS MEASURED	INDEX DESCRIPTION	INTERVIEW CONTACT AND TITLE (IF APPLICABLE)	CITATION OR WEBSITE
Community Fluoridation Compliance Index	Public Health Reports (Journal)	Level of compliance with respect to regulatory requirement for fluoridation levels	This article describes the development of a Community Fluoridation Compliance Index (CFCI), which was designed to provide retrospective and prospective information on water systems in complying with local and State standards. Data from 50 water systems in Ohio and 50 systems in Illinois were used to create the CFCI, which calculates how close the water system is to achieving optimal fluoridation concentrations over a given time period.		Johansson, Robert C. and Cattaneo, Andrea. "Indices for Working Land Conservation: Form Affects Function." *Review of Agricultural Economics*, Vol. 28, No. 5, 2006. pp. 567-584. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1477864/pdf/pubhealthrep00176-0069.pdf
Disclosure Compliance Index	Accounting and Business Research (Journal)	Compliance with financial disclosure requirements	This paper empirically examines the level of compliance with disclosure requirements mandated by 14 national accounting standards for a large sample of companies in India, Pakistan and Bangladesh. The 14 standards were developed into 131 disclosure compliance requirements. Companies were given a score of zero (noncompliant) or one (compliant) for each requirement. The Total Compliance Index is calculated as the ratio of the total disclosure score to the number of applicable requirements.		Ali, Muhammad Jahangir. "Disclosure compliance with national accounting standards by listed companies in South Asia." Accounting and Business Research, Vol. 34, No. 3, 2004. pp. 183-199.
Compliance Rate Template	Environmental Compliance Consortium	Facility compliance	This article describes the Environmental Compliance Consortium's (ECC) Compliance Rate Template, a spreadsheet that was developed to help states report their compliance levels. Presents the following information in addition to the compliance rate percentage: • The percentage of facilities where compliance was monitored; • When available, the reasons for monitoring compliance; and • Other key characteristics important to interpreting compliance rate information, such as whether statistically valid methods were used to the select the monitored sites and whether the inspections were pre-announced.		Shewmake, Tiffin. "Calculating and Communicating Environmental Compliance Rates." *ECOStates Journal*, Spring 2003. pp. 23-27. Shewmake, Tiffin. "Using Compliance Rates to Manage." *ECOStates Journal*, Fall 2004. pp. 17-22.

INDEX NAME	AGENCY/ ORGANIZATION	WHAT IS MEASURED	INDEX DESCRIPTION	INTERVIEW CONTACT AND TITLE (IF APPLICABLE)	CITATION OR WEBSITE
Environmental Performance Index – Portuguese Military	Business Strategy and the Environment (Journal)	Self-assessed environmental performance	The main goal of this research was to assess the Portuguese military sector's environmental performance, through a simplified and useful manner, through an index in a simplified and useful manner. Survey questions were paired down and then scaled from one to zero. The scores were then aggregated using equal weights. A second goal was to measure the units' self-assessment against the environmental profile of the military sector (as evaluated through the index) to determine how well they match.		Ramos, Tomas B. and de Melo, Joao Joanaz. "Developing and Implementing an Environmental Performance Index for the Portuguese Military." *Business Strateg and the Environment*, Vol. 15, 2006. pp. 71-86.
Environmental Performance Index	Yale Center for Environmental Law & Policy and Columbia University Center for International Earth Science Information Network	Global environmental performance	Developed an Environmental Performance Index (EPI), which is a global index of overall environmental performance. The EPI focuses on two overarching environmental objectives: • Reducing environmental stresses to human health; and • Promoting ecosystem vitality and sound natural resource management.		"2008 Environmental Performance Index." Yale Center for Environmental Law & Policy and Columbia University Center for International Earth Science Information Network (CIESIN). Available at: http://epi.yale.edu.
Environmental Results Program (ERP)	US EPA	Facility performance levels on selected indicators of regulatory compliance and best practices	ERP programs involve selecting a set of compliance and beyond compliance indicators for which a statistically-based random sample of facilities are measured at baseline and follow-up inspections. Inspection results are compared and reported in a number of ways, including by environmental media.		http://www.epa.gov/erp/

INDEX NAME	AGENCY/ ORGANIZATION	WHAT IS MEASURED	INDEX DESCRIPTION	INTERVIEW CONTACT AND TITLE (IF APPLICABLE)	CITATION OR WEBSITE
IRS Compliance Index	Internal Revenue Service	Voluntary filing compliance; Reporting compliance; and Payment compliance	IRS started the National Research Program to better assess national taxpayer compliance levels. Compliance is measured in three ways: voluntary filing, reporting, and payment compliance. Statistical random stratified samples of tax returns are audited to gather data, with weighted average scores used for aggregation.	Bob Brown, Office of Research, Analysis and Statistics	http://www.irs.gov/privacy/article/0,,id=139179,00.html IRS National Research Program. "Challenges Associated with Collection Compliance Data." IRS Research Conference, 2002. Brown, Robert E. and Mazur, Mark J. "IRS's Comprehensive Approach to Compliance Measurement." IRS, June 2003. "National Research Program: Early Results & Future Effort." IRS Office of Research, Analysis, and Statistics Presentation, June 15, 2003.
Massachusetts Capital Assets Management Deferred Maintenance Process	MA Department of Capital Assets Management (MA DCAM)	Building condition and priority for specific maintenance projects	MA DCAM maintains a management information system to identify deficiencies in approximately 5,000 state-owned building. The system is based on an initial condition assessment completed in 2000. MA uses a scoring system to allocate annual deferred maintenance across all 500 buildings. The scoring system is based on a series of indicators based on the initial assessment and updates provided by the building managers and confirmed by staff from MA DCAM.	Hope Davis, Director of Office of Facilities Maintenance and Management	http://www.mass.gov/?pageID=afsubtopic&L=3&L0=Home&L1=Property+Management+%26+Construction&L2=Facilities+Management+%26+Maintenance&sid=Eoaf FY 10 Deferred Maintenance Process, Memo from David Perini, Commissioner of Division of Capital Asset Management. March 23, 2009.
Ontario Compliance Index	Ontario Ministry of the Environment	Relative compliance levels	Developed a compliance index as a measure of relative compliance levels, based on types of legislative violations. It provides a performance measure for facility, sector, and media/program compliance levels. The index is a weighted sum of facility scores on applicable regulatory requirements.	Rajeev Narang, Environmental Program Analyst, Sector Compliance Branch	Ontario Ministry of the Environment. *Compliance Index(CI): Draft Report.* June 3, 2008. Narang, Rajeev, Ontario Ministry of the Environment. "Compliance Index – Environmental Compliance/Performance Measure." Presented at the National Environmental Partnership Summit, Baltimore, MD. May 20, 2008.

INDEX NAME	AGENCY/ ORGANIZATION	WHAT IS MEASURED	INDEX DESCRIPTION	INTERVIEW CONTACT AND TITLE (IF APPLICABLE)	CITATION OR WEBSITE
Operator Pollution and Risk Appraisal (OPRA)	UK Environment Agency	Potential risk associated with facility operations	The EA developed this rating tool as part of its integrated permitting process. The tool is used to produce a banded score based on a facility's complexity, emissions and inputs, location, compliance and operator performance. The ratings then determine permit fees and also inform enforcement and monitoring priorities for the agency.	David Pugh Better Regulation Policy Manager, United Kingdom Environment Agency	http://environment-agency.gov.uk/opra UK Environment Agency. "Spotlight on business – 10 years of improving the environment." July 2008. (see page 20 and pages 26-33) http://publications.environment-agency.gov.uk/pdf/GEHO0708BOFX-E-E.pdf?lang=_e U.S. EPA, National Center for Environmental Innovation. "An In-Depth Look at the United Kingdom Integrated Permitting System." July 2008. http://www.epa.gov/permits/IntPermittingRpt.pdf
Operator Safety Index (OSI)	Department of the Interior's Mineral Management Service (DOI MMS)	Operator compliance with federal safety regulations and accidents	The DOI MMS created this index to measure operator compliance and accidents at offshore oil and gas facilities. The MMS conducts facility inspections and measures instances of noncompliance and gathers information about any accidents at the facility (e.g., property damage, amount of spillage, injuries). Instances on noncompliance are weighted by the type of enforcement action (e.g., warning, civil penalty) and normalized by the size of the operations. Information about the accidents is combined with a "severity value" to create an accident severity score. These two factors are combined to create the OSI. The OSI is used to nominate safety award winners and to inform annual performance review meetings with facilities, operator disqualification, and inspection strategies.	Doug Slitor, Chief, Safety and Enforcement Branch	Slitor, Doug. "Measuring Safety and Compliance in the U.S. Offshore Oil and Gas Industry.: *Society of Petroleum Engineers*. Paper prepared for presentation at the SPE International Conference on Health, Safety, and the Environment in Oil and Gas Exploration and Production held in Stavanger, Norway, June 26-28, 2000. SPE 61155.

CHAPTER 2 | REVIEW OF INDEX DEVELOPMENT APPROACHES

INTRODUCTION

Developing a compliance measurement and reporting approach requires regulators to ask several questions. For example, what is going to be measured? How will the data be collected? How will the results be reported? At one extreme, regulators can measure *all* the compliance-related indicators, and report compliance as an all-or-nothing measure. In other words, a facility would be considered "in compliance" only if it met *all* of its regulatory requirements (or all of a pre-specified subset of the requirements). This approach would provide a regulator with a count or percentage of facilities in complete compliance (i.e., what we have called so far a compliance "rate"). In this chapter we explore a range of options in designing an alternative approach for assessing compliance based on using indexes, focusing on the following key elements:

- Indicator selection;
- Data collection; and
- Reporting.

For each of these elements, we discuss a range of possible approaches, as well as the trade-offs that accompany different choices within those ranges. For example, when thinking about indicator selection, a regulator must decide whether to include all indicators of compliance, or a subset of indicators that reflect either priority areas or practices that are considered fundamental to minimizing environmental impacts. For each element, we cite specific examples from the case studies we have compiled (please see Appendix A for more detailed descriptions of these case studies).

INDICATOR SELECTION

While ideally regulators may want to measure every indicator of compliance for every facility, this is often not realistic given budgetary and time constraints, especially in sectors with a large number of facilities and complex regulatory requirements. Instead, regulators can choose a subset of the complete list of regulatory requirements in order to focus on the most important aspects of concern and more efficiently allocate their resources. In this process, regulators face a trade-off between comprehensive compliance measurement and the capacity to implement feasible data collection approaches.

When selecting indicators, regulators should consider:

- Relevancy to program goals,
- Feasibility of gathering data, and

- Availability of sufficient data to measure compliance meaningfully.

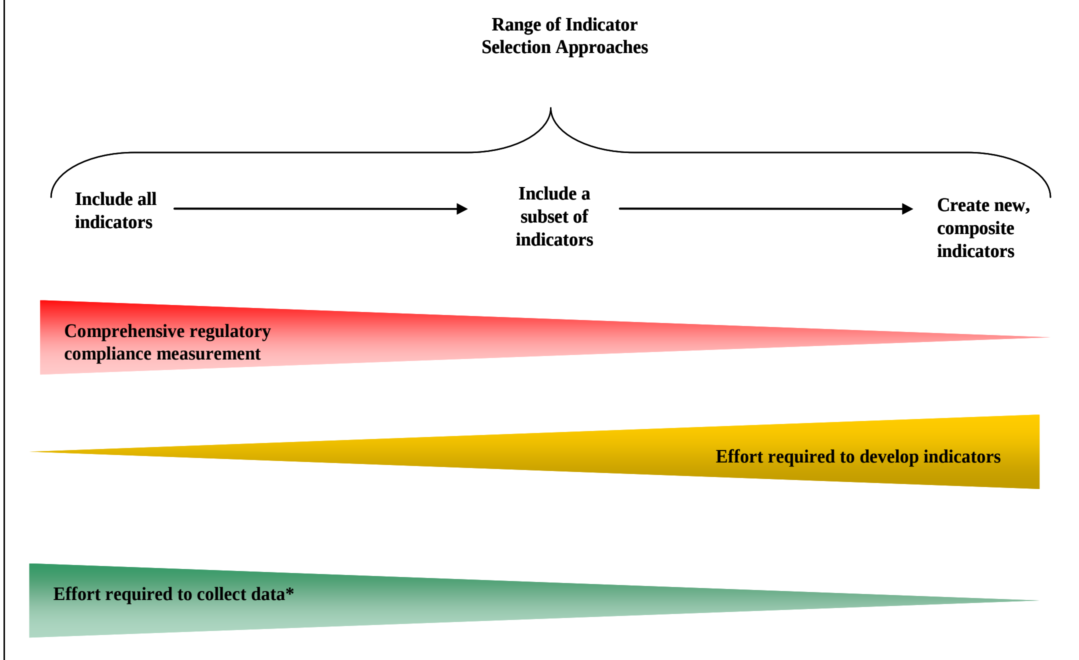
Regulators should approach the process of creating an index of indicators with an eye toward the intended use of the index, choosing indicators that accurately reflect priority environmental concerns or regulatory requirements. In other words, the “core” indicators of compliance for a given regulation should be measured. Regulators should also consider whether the available time, resources and technologies will be sufficient to collect the necessary data for the indicators. A final consideration involves ensuring that the selected indicators will provide enough information to determine facility compliance, and understand facility behaviors.

Regulators should also consider whether they are interested in measuring “leading” indicators that inform the *potential* performance, or risk, of a facility, or “lagging” indicators that reveal *actual* performance of a facility. For example, a leading indicator could be whether or not the facility properly stores its hazardous waste. This indicator is leading because it helps the regulator understand the facility’s potential for a waste spill; if the facility is properly storing its waste, it is less likely to have an accident. On the other hand, a lagging indicator could be whether the facility has had any hazardous waste spills. This indicator is lagging because it tells the regulator something about the actual performance of the facility; whether or not it has had spills. Both kinds of indicators provide valuable information to the regulator about the facility’s performance; when choosing indicators to include in an index, regulators should consider what kind of performance they are trying to measure and structure the set of indicators accordingly.

The U.S. Department of Interior’s Minerals Management Service (MMS) has created an index that combines both leading and lagging indicators because it wants to measure both performance and potential safety risks. This index, called the Operator’s Safety Index, is comprised of two components: incidents of noncompliance, and accident severity. The instances of noncompliance component is essentially a “leading” indicator of operations; in other words, it represents how risky the operations of a facility are in terms of how well it is meeting safety and operational requirements). The accident severity component serves as a “lagging” indicator because it captures the actual performance of a facility with respect to safety. In combination, these indicators help the agency understand both the past performance of a facility, and the potential risk it poses to safety.

Exhibit 2-1 describes a spectrum of indicator selection approaches regulators face: from including all indicators, to including a subset of existing indicators, to creating new composite indicators (i.e., new indicators that combine aspects of existing indicators). It shows the trade-offs along this spectrum in terms of the level of regulatory compliance measurement achieved, the level of effort required to develop indicators, and the level of effort required to collect compliance data. Choosing to include all indicators in an index ensures comprehensive measurement of compliance, but requires more effort to collect data at each facility. As shown on the left end of the spectrum in the exhibit, this approach should generally require fewer resources to develop, as it is straightforward to identify the full list of indicators relating to regulatory requirements and other priorities.

EXHIBIT 2-1: SPECTRUM OF INDICATOR SELECTION APPROACHES



*It could actually require more effort to collect inspection data for subsets or new set of indicators, if inspectors must be re-trained.

Ontario's Ministry of the Environment (MOE) chose this approach when developing its compliance index. The MOE decided that all applicable legislative provisions should be incorporated into its index. The Ministry was already collecting compliance information about all of the facilities' applicable legislative requirements; therefore, the compilation of this information into an index did not require changes to how the information is collected (e.g., either the number of or process for inspections). All that had to be changed is how that information is compiled and reported. Furthermore, since all of the indicators in the index were based on data already being collected during inspections, MOE could continue to use its current information management systems.

In the middle of the spectrum, is the option to include a subset of indicators. Environmental Results Programs (ERPs) generally follow this approach by developing a set of indicators from a full list of regulatory requirements. Often, ERPs also include new indicators that represent best practices for the target sector. This approach allows the program to focus on areas of concern, and efficiently measure both sector compliance and environmental performance. While this approach requires a larger investment in resources to develop indicators than using the full list of regulatory requirements, a shorter, more straightforward inspection process often saves agency resources. This approach also sacrifices a degree of comprehensiveness, as all indicators of compliance are not included; however, the indicators that measure best practices can provide additional information about overall facility performance.

Finally, on the far right end of the spectrum, there is the option to create a new composite set of indicators. With this approach, regulators would develop a new, shorter list of indicators that incorporates the requirements of the full list of indicators. In other words, the new indicators would streamline the data collection process, but would capture the primary measures of compliance of the full list of indicators. Yale/Columbia used this approach to develop the Environmental Performance Index. Using data from several countries and several sources, researchers developed 25 indicators that represent "composite" topics, such as drinking water quality and indoor air pollution. These indicators were chosen to represent a broad range of environmental challenges. A significant amount of effort was required to complete this effort, including consultation with several other organizations, subject-area experts, statisticians, and policymakers from around the world. This process was critical for combining and comparing data from such disparate sources.

DATA COLLECTION

Once a set of indicators has been developed, regulators must decide on a data collection approach. Data collection in the compliance measurement context most often means facility inspections; however, regulators can also rely on self-reported data, such as surveys or questionnaires. Collecting compliance data without the use of inspections forces regulators to forfeit a level of verification (i.e., they must rely on the integrity and capability of the facility to provide accurate data).[9] In this section we focus on the inspection approach to data collection because it is the most common method used in the compliance measurement context.

Regulators may also decide to construct an index using existing data that they have already collected. The motivation for this decision may be the desire to spend fewer resources, or that the data already collected sufficiently provides the information the regulator needs to understand compliance of the sector. Therefore, data collection may not necessarily involve collecting new data, but it instead could be based on sorting through and reexamining existing data. However, in this section, we focus primarily on the collection of new data.

Similar to the decision about indicator selection, when designing an inspection approach, regulators must choose between inspecting *all* facilities in a sector, or a subset of those facilities. Again, regulators would ideally like to inspect every facility every year; however, they often lack the resources to do so. The decision-making criteria used in this process can dramatically affect the calculated compliance rate and its interpretation.[10] For example, if only facilities that have previously been subject to enforcement actions are targeted for inspections, the resulting compliance rate could be applied only to those facilities.

The inspection approach adopted should also be tailored to the characteristics of the sector in question. For example, some OECA guidance instructs regulators to inspect every facility in the sector at regular intervals if the sector is composed of primarily major

---

9 "Principles of Environmental Compliance and Enforcement Handbook". International Network for Environmental Compliance and Enforcement (INECE). April, 2009.

10 Shewmake, Tiffin. "Calculating and Communicating Environmental Compliance Rates." ECOStates Journal, Spring 2003. pp. 23-27.

sources.[11] If, on the other hand, the sector is composed of a large number of area sources, inspecting every facility will most likely be impractical.

We focus on three possible data collection approaches here:

- Inspect the entire universe of facilities;
- Select a statistically based random sample of facilities from the universe or from a targeted subgroup of facilities; or
- Inspect a targeted group of facilities.

Exhibit 2-2 summarizes the advantages and disadvantages of each of these approaches.

There are several circumstances in which each of these approaches is appropriate. For example, with funding from a State Innovation Grant (SIG) from EPA, the Connecticut Department of Environmental Protection undertook a compliance measurement effort in which it inspected every SQG facility in the state over a four year period. While data from each inspection year was not compiled into a comprehensive measure of compliance (i.e., each year's measure of compliance was calculated separately), the agency wanted to inspect every facility by the end of the four year process. In this case, inspecting every facility met the agency's goals, and was feasible as a result of the EPA grant money.[12]

In a study of water system compliance with community fluoridation level regulations across two states, it was not feasible to collect data from all of the water systems in the states.[13] Therefore, to enable extrapolation of data from the selected water systems to the entire state-wide populations of systems, the study authors took a statistical random sample of water systems. This allowed the authors to undertake a manageable data collection effort while simultaneously providing state-wide information.

[11] For example, the EPA's Compliance Monitoring Strategy for the Clean Air Act states that: "A Full Compliance Evaluation should be conducted, at a minimum, every two years at all Title V major sources except those classified as mega-sites. For mega-sites, a Full Compliance Evaluation should be conducted, at a minimum, once every three years". U.S. EPA, "Clean Air Act: Stationary Source Compliance Monitoring Strategy", April 2001, page 9. Available at: http://www.epa.gov/compliance/resources/policies/monitoring/cmspolicy.pdf. (Last viewed November 12, 2009).

[12] Personal communication with Susan Zampaglione, CT Department of Environmental Protection, October 23, 2009. Note that CT initiated this effort under its own SIG. After the Common Measures project was funded, CT participated in this larger effort using one year of this SQG performance data.

[13] Kuthy, Raymond A. "Use of a Compliance Index for Community Fluoridation." Public Health Reports, Vol. 102, No. 4, July-August, 1987. pp. 415-420.

EXHIBIT 2-2: SUMMARY OF DATA COLLECTION APPROACHES

APPROACH	ADVANTAGES	DISADVANTAGES
Entire universe of facilities	• Confidence that every facility has been inspected. • Satisfy stakeholders that data provide an accurate picture of compliance. • Meet OECA guidance to inspect every facility (for major sources).	• Can require more resources to complete, especially in sectors with larger numbers of facilities. • Need to have identified every facility in the universe.
Statistically random sample of universe or targeted subgroup of facilities	• Achieve the same confidence in understanding the performance of the entire universe of facilities. • Spend fewer resources to achieve the same level of confidence in results.	• Must be able to accurately characterize the universe of facilities, i.e., to know the number of facilities. • Need technical expertise to conduct sampling calculations. • Due to limited resources, often must sacrifice targeted ("for cause") inspections for random inspections. • Can be difficult to explain statistical concepts to a broader audience (i.e., explain how results from a sample represent the entire universe). • When selecting from a subgroup, results will not apply to the full universe.
Targeted facilities	• Confident that data provide understanding of the performance of facilities of concern. • Generally easier to identify facilities (than entire universe). • Spend fewer resources because fewer inspections conducted (for sectors with large numbers of facilities).	• Not capturing performance data about the majority of the sector (including facilities that are performing well). • Not likely to produce a complete picture of "typical" performance levels. • Results not statistically valid and therefore cannot be generalized to the full universe of facilities.

In some cases, regulators may want to target a subgroup of the universe of facilities, but still be able to make statistical inferences about that subgroup without inspecting every facility in it. This was the case for the IRS in its compliance measurement program. Measuring compliance of all of the taxpayers in the country is clearly not feasible. To solve this problem, the IRS takes a statistical random sample of each type of taxpayer for every tax year (e.g., individual taxpayers, small businesses, etc.). This allows the agency to conduct tax return audits on a feasible number of returns, and still make inferences about the entire population of taxpayers based on the results of those audits.

We did not find any examples of indexes that were calculated based on data collected from a targeted group of facilities; however, this approach seems a likely extension of the continuum of data collection approaches. An index could be constructed to characterize the compliance of a targeted group of facilities, such as a set of particularly complex facilities with potentially large environmental impacts. This approach would characterize the performance of these specific facilities of concern; however, it would not provide information about the performance of the sector or universe as a whole.

REPORTING

The final step in creating a compliance index is deciding how to report the results. Reporting results in a thoughtful way can help agencies:

- Monitor performance;
- Analyze performance of subsets of regulatory requirements;
- Review effectiveness of specific programs;
- Report to external audiences;
- Assess and adapt indicators; [14]
- Find anomalies such as low compliance levels on relatively achievable requirements;
- Search for correlations among different types of compliance violations; and
- Measure progress and identify persistent problems.[15]

We propose three primary categories to consider for compliance measurement reporting: measure of compliance, weighting of indicators to reflect potential impacts or policy priorities, and grouping of results.

Measure of Compliance

The most fundamental result of a compliance measurement effort is whether or not facilities are in compliance with regulations. There are several ways to report compliance, including:

- **Overall sector percent of compliance.** When measured for all requirements (i.e., 75 percent of facilities are meeting all of their requirements), this is essentially a compliance rate.

[14] "Performance Measurement Guidance for Compliance and Enforcement Practitioners: Second Edition". International Network Environmental Compliance and Enforcement (INECE), Expert Working Group on Enforcement and Compliance Indicators. April, 2008.

[15] Metzenbaum, Shelley. "Compliance and Deterrence Research Project: Measuring Compliance Assistance Outcomes: State of Science and Practice White Paper". Prepared for U.S. EPA Office of Enforcement and Compliance Assurance. December 6, 2007.

- **Aggregate facility percent of compliance.** A regulator may want to measure the performance of an entire sector (or subgroup within a sector) in a way other than the overall percent of compliance. The aggregate facility percent of compliance essentially creates one large composite facility, and reports the percent of compliance with all requirements. While this measure still reports one overall number for the sector or group of facilities, it provides more information than the overall sector percent of compliance because it tells the regulator how well, as a whole, that group is performing (i.e., how many of the requirements the facilities are meeting). While the overall sector percent of compliance reports the percent of *facilities* that are in full compliance; the aggregate percent of compliance reports the percent of *requirements* that the facilities are meeting.

- **Facility-specific percent of compliance.** A facility-specific percent of compliance measures the overall performance of the facility and can be measured for all or a subset of requirements. For example, a facility can be measured as meeting 75 percent of its total requirements, or as meeting 75 percent of its air emission requirements. This measure captures more subtle information about the facility's performance because it tells regulators how close or far the facility is to complete compliance. For example, compare two facilities: one is 20 percent in compliance and the other is 80 percent in compliance. With a traditional compliance rate, both of these facilities would be "out" of compliance, but with a facility-specific percent of compliance, regulators can clearly see that the second facility is "closer" to full compliance.

- **Indicator-specific percent of compliance.** Similar to the facility-specific percent of compliance, regulators can calculate percent of compliance specific to a single indicator or group of indicators (e.g., 75 percent of facilities are in compliance with the requirement to label their hazardous waste containers). This approach can be useful when a regulator wants to identify regulatory requirements posing special compliance challenges to the sector, e.g., to focus on and address noncompliance with especially high risk indicators.

Regulators can also apply a scoring system to create a scaled index. This approach can help translate numerical results into meaningful values. For example, facilities could be assigned a category label based on the percent of total compliance (e.g., poor, good, excellent). Instead of reporting that a facility (or group of facilities) is meeting 23 percent of its regulatory requirements, regulators can label the facility (or group of facilities) "poor performers." While sacrificing a level of precision, this kind of labeling or naming may add meaning to the index for the general public and other stakeholders.

This scoring approach can also help communicate efforts to track progress; compare results across programs, types of regulatory requirements, or over time; and create benchmarks or goals. For example, after discovering that a significant number of the facilities in a sector are "poor performers," regulators could set a goal of half of those facilities improving to "good performers" following the next round of compliance assistance and enforcement actions. Compared to a numerical target, this kind of goal is more easily communicated to stakeholders and the general public.

In a study to assess the environmental performance of the Portuguese military, the authors scored each indicator and aggregated them into an overall score of environmental performance on a scale from zero to one. Then the authors translated the score to a five-point scale to categorize performance on a scale from very poor to excellent. This scaled score could then be used to compare performance over different time periods or across subgroups of the military. Exhibit 2-3 presents the scaled scoring system used in this index.

EXHIBIT 2-3: PORTUGUESE MILITARY ENVIRONMENTAL PERFORMANCE INDEX SCORING SYSTEM

INDEX SCORE	ENVIRONMENTAL PERFORMANCE CATEGORY
0.0 – 0.20	Very Poor
0.21 – 0.40	Poor
0.41 – 0.60	Medium
0.61- 0.80	Good
0.81 – 1.0	Excellent

The United Kingdom's Environment Agency (EA) uses an operational risk appraisal scheme (Opra) to assess the potential risk associated with facilities and help target agency resources to more effectively meet its objectives of environmental protection. A facility is scored in five attributes (complexity, emissions and inputs, location, operator performance, and compliance), and these scores are then used to create a "banded profile" on a five-point scale (from A to E) for each attribute. EA uses these scores to establish permitting and annual fees for facilities, plan inspection activities, track progress, set goals, and also to communicate results to the public. Exhibit 2-4 below provides an example of a graphic from the EA's report "Spotlight on Business: 10 years of improving the environment" that is used to communicate progress to the public. The graphic presents the overall operator performance of industry from 2005 to 2007, and clearly shows the distribution of scores on the five-point scale.[16]

EXHIBIT 2-4: EXAMPLE OPRA RESULTS GRAPHIC – TRACKING PROGRESS

[16] UK Environment Agency. "Spotlight on business – 10 years of improving the environment." July 2008.

Weighting of Indicators

Indicators can be weighted to emphasize (using larger weights) or minimize (using smaller weights) their impact on the overall measure of compliance. In our research, some common criteria for weighting included: level of risk; amount of environmental impact; and importance of the regulatory requirement. While there are many different criteria that can be used to weight indicators, it is important to consider the how the alternative weighting schemes might affect the “message” of an index. It is also important to be cognizant of how adding weights to indicators transforms the original results.

Weighting can be an efficient and effective way to characterize the performance of a facility (or group of facilities) based on a set of criteria. Given a large set of indicators that represent a long list of regulatory requirements, inevitably there will be indicators that capture varying levels of environmental impacts. If the indicators that pose a higher environmental risk are weighted more heavily than those that are less risky, regulators can potentially identify facilities that are engaging in more risky noncompliant activities.

Exhibit 2-5 below provides an illustration of the impact of weighting on index scores. Under scenario one, all four indicators are weighted evenly and the two facilities have identical noncompliance scores. Once differential weights are applied to the indicators reflecting their relative potential impacts, the noncompliance score for the first facility becomes much higher than the second. This example demonstrates how weighting can allow a regulator to differentiate between the extent of compliance for two facilities even though the facilities have the same number of violations.

EXHIBIT 2-5: IMPACT OF WEIGHTING ON NON-COMPLIANCE SCORES

INDICATOR NUMBER	WEIGHT	FACILITY 1		FACILITY 2	
		COMPLIANT?	NONCOMPLIANCE SCORE	COMPLIANT?	NONCOMPLIANCE SCORE
SCENARIO 1: EQUAL WEIGHTS APPLIED TO ALL INDICATORS					
1	0.25	No=1	0.25	Yes=0	0.0
2	0.25	No=1	0.25	Yes=0	0.0
3	0.25	Yes=0	0.0	No=1	0.25
4	0.25	Yes =0	0.0	No=1	0.25
		Total Score:	**0.50**	**Total Score:**	**0.50**
SCENARIO 2: DIFFERENT WEIGHTS APPLIED TO EACH INDICATOR					
1	0.40	No=1	0.40	Yes=0	0.0
2	0.40	No=1	0.40	Yes=0	0.0
3	0.10	Yes=0	0.0	No=1	0.10
4	0.10	Yes=0	0.0	No=1	0.10
		Total Score:	**0.80**	**Total Score:**	**0.20**

To calculate its academic performance index (API), the California Department of Education assigns weights to six degrees of academic performance. The index is calculated by converting a student's performance on state-wide assessments across multiple content areas into performance levels on a scale (see Exhibit 2-6 below). Each performance level is then assigned a weighted score. The weights progressively increase with higher performance levels, to essentially reward a school for having more students performing at a higher level. The percentage of students at each level is then multiplied by the respective weight, and the resulting values for each level are summed to produce the school's overall performance index score. These scores can then be compared across schools, and can also be used to track performance from year to year.

EXHIBIT 2-6: CALIFORNIA API WEIGHTS

PERFORMANCE LEVEL	% OF STUDENTS AT LEVEL	WEIGHT	SCORE
Untested	5	0.0	0.0
Limited	20	0.3	6.0
Basic	25	0.6	15.0
Proficient	35	1.0	35.0
Accelerated	5	1.1	5.5
Advanced	10	1.2	12.0
Performance Index Score			**73.5**

Ontario's MOE assigned weights to its indicators (different types of violations) based on the level of potential impact on the environment and human health, with higher weights representing a larger impact. This means that a facility and/or sector could achieve a similarly high score by having either a large number of low impact violations or a small number of high impact violations. Assigning high impact indicators a larger weight ensures that a facility will not appear to have high compliance (low score) when it has high impact violations. At this time, MOE uses a weighting scale with three values (1, 2, or 4). The agency is considering further refinements to its weighting scheme to better capture differences in the potential impacts of the various types of violations.

We also note that in one of the case study indexes, the environmental performance index for the Portuguese military, the authors constructed the index in a way that allows for a weighting scheme to be applied across a set of 18 indicators; however they chose not to use it in calculating their index. They acknowledge that "in order to consider the relative importance of each different indicator, the … index allows for weighting, but for the Portuguese military the … index was computed with equal weights for each indicator."[17]

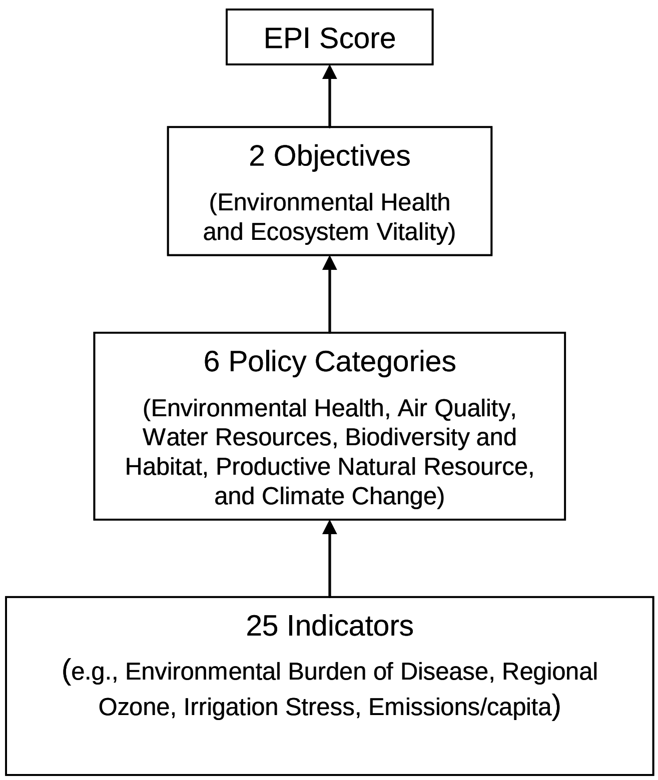
Grouping of Results

Regardless of the data collection approach used, it is possible to group results in a number of ways. The extent to which indicators should be grouped and aggregated depends on

[17] Ramos, Tomas B. and de Melo, Joao Joanaz. "Developing and Implementing an Environmental Performance Index for the Portuguese Military." Business Strategy and the Environment, Vol. 15, 2006, page 76.

who is going to be using the results and for what purpose.[18] Individual indicators can be grouped, for example, by environmental media, by sector, or by type of regulatory requirement. In developing their Environmental Performance Index (EPI), researchers from Yale and Columbia measure 25 indicators that are aggregated into six policy categories of interest. These six policy categories are further grouped into two overall environmental objectives: environmental health and ecosystem vitality. Finally, the EPI is calculated as the arithmetic mean of the environmental health and ecosystem vitality scores. See Exhibit 2-7 below for an illustration of this process. This approach is effective for distilling a large amount of data first into 25 indicators, then six groups, then two policy objectives, and finally an overall score. Furthermore, this hierarchy of indicators provides considerable flexibility in communicating index results to stakeholders.

EXHIBIT 2-7: YALE/COLUMBIA ENVIRONMENTAL PERFORMANCE INDEX CONSTRUCTION



In some cases, regulators may want to examine index scores by environmental media to discover if there are any significant differences. Exhibit 2-8 below shows an example from ERP where indicators are coded in a way that allows for easy reporting by environmental media (e.g., air, water, hazardous waste). The exhibit provides the number of indicators in each category, as well as the number of indicators improving, worsening, or staying the same. Because ERP uses statistical methods to calculate changes in

[18] Hammond, Allen et al. "Environmental Indicators: A Systematic Approach to Measuring and Reporting on Environmental Policy Performance in the Context of Sustainable Development." World Resources Institute, May 1995.

performance levels, the table also presents the number of indicators improving and worsening that are statistically significant.

EXHIBIT 2-8: ERP RESULTS GROUPED BY ENVIRONMENTAL MEDIUM

ENVIRONMENTAL MEDIUM	# OF INDICATORS	# IMPROVING (# SIGNIFICANT)	# WORSENING (# SIGNIFICANT)	# NO CHANGE
Air Emissions	14	9 (6)	3 (0)	2
Waste Management	19	16 (9)	2 (0)	1
Water Discharge	9	5 (4)	4 (0)	0
Worker Health and Safety[(b)]	5	3 (3)	2 (0)	1
Overall	**47**	**33 (22)**	**11 (0)**	**4**

In all data collection scenarios, inspectors will collect data about whether or not the facility is meeting each indicator. This data can then be grouped at the facility level, or by indicator (across all facilities). Among the case study indexes we examined, the majority report results at the facility level. In other words, the indexes report the total "score" for each facility (or group of facilities). For example, in a study of compliance with accounting disclosure requirements in South Asia, the authors calculated an index score for each company in their sample; the scores from each company can then be compared.[19] Three of the case study indexes (ERPs, the UK's Opra, and the Common Measures Project) report results at the indicator level; results are reported as the total number (or percentage) of facilities achieving a single indicator or group of indicators (see Exhibit 2-8 above for an example of ERP results grouped by environmental medium). This approach of grouping results by indicator (or group of indicators) can be useful in examining differences across the facilities with respect to certain categories of concern.

SUMMARY OF INDEX CONSTRUCTION APPROACHES

While it is important to consider each of these factors individually when determining how to construct an index, it is also important to consider them as a whole. Regulators may want to think of index construction as a step-by-step thought process, in which the particular circumstances of their situation helps inform the design of the final index. As an example, consider a regulatory agency that has an interest in characterizing the performance of a specific set of facilities within the universe of facilities under its jurisdiction: facilities that have received enforcement actions. This agency wants to know how these facilities are performing with respect to a certain number of key indicators. Furthermore, it does not have the time or money to inspect all of these facilities, and its main concern is the group's performance as a whole. Exhibit 2-9 below describes a step-by-step process that this agency could follow as it considers how to build an index to best meet its needs.

[19] Ali, Muhammad Jahangir, Ahmed, Kamran, and Henry, David. "Disclosure Compliance with National Accounting Standards by Listed Companies in South Asia." *Accounting and Business Research*, Vol. 34, No. 3, 2004. pp. 183-199.

EXHIBIT 2-9: SUMMARY OF INDEX CONSTRUCTION APPROACHES

1 – Indicator

Question: What indicators should be included in the index?

All indicators | $\boxed{\text{Subset of indicators}}$ | New, composite indicators

Why? Agency has limited resources, and is trying to capture a few key indicators of compliance.

2 – Data Collection

Question: How will the data be collected/obtained?

Entire universe | $\boxed{\text{Statistically random sample}}$ | Targeted

Why? Agency does not have sufficient resources to inspect all facilities. It also needs to gather information about a subgroup of facilities (those that have had enforcement actions).

3 – Measure of Compliance

Question: How will the index be reported?

Overall sector | $\boxed{\text{Aggregate facility}}$ | Facility-specific | Indicator-specific

Why? Agency wants to know about the performance of this subset of facilities as a whole.

4 - Weighting of indicators

Question: Should any of the indicators be weighted more heavily than others?

Yes | $\boxed{\text{No}}$

Why? Agency is already selecting a subset of indicators that highlight the issues the it cares about.

5 - Grouping of results

Question: Should the results be grouped? If so, how?

$\boxed{\text{Yes, by statute}}$ | No

Why? There are a number of statutes the agency is measuring, and it wants to see if there are performance differences across the statutes.

Reporting Options

CHAPTER 3 | FINDINGS AND CONSIDERATIONS

INTRODUCTION

In this chapter, we present a series of considerations in developing indexes based on our review of the literature and the 13 case study indexes that we identified and explored as part of this effort. The considerations fall into four major categories: use of the index, selection of indicators, data collection, and weighting of indicators. We describe each of these categories and illustrate them with examples from the case study indexes. In addition, we include an exhibit at the end of this chapter that highlights specific key lessons from these indexes.

USE OF INDEX

The ultimate use of a compliance index is a key factor in determining how it should be developed. One important dimension of this factor is whether the index is intended to complement or replace an aspect of traditional compliance measurement. To the extent an index serves the former purpose, there will be greater flexibility in how it can be designed (e.g., selection of indicators, assigning of weights and interpretation of scores). When designing an index with the goal of replacing existing or traditional reporting approaches, a regulator will need to demonstrate that it produces comparable results in terms of covering the same indicators and providing an equivalent level of confidence in determining compliance.

A similar distinction applies to indexes that are intended to measure performance as opposed to those that strictly measure compliance. Among the case study indexes included in this report, we found three that addressed both performance of non-regulated, or beyond compliance activities and compliance with regulatory requirements (ERP, the Common Measures Project, and Opra), four that measure performance (the Environmental Performance Index developed by Yale and Columbia, the Environmental Performance Index for the Portuguese Military, the Academic Performance Indicator, and the Massachusetts Capital Assets Management Deferred Maintenance Process), and six that measure compliance with regulatory requirements (the Disclosure Compliance Index, Community Fluoridation Compliance Index, the ECC Compliance Rate Template, the IRS Compliance Index, the Ontario Compliance Index, and the DOI MMS Operator's Safety Index). In general, we found that development of those indexes focused on strict compliance measurement was driven by the specific requirements being measured. Developers of indexes for measuring general areas of performance tended to have greater flexibility in defining indicators and in how policy priorities were incorporated into both the indicators and weighting schemes.

SELECTION OF INDICATORS

Many of the case study indexes we examined illustrate the importance of indicator selection. As noted in Chapter 2, there are a series of criteria for selecting indicators commonly identified in the literature, including: relevancy to program goals, feasibility of gathering data, and availability of sufficient data to measure compliance meaningfully.

We have identified the following specific considerations concerning the selection of indicators:

- **Selection of all or a subset of requirements.** Case study indexes differed in whether they focused on all or subsets of regulatory requirements. For example, the Common Measures Project included a detailed examination of SQG requirements to identify those that were common to and considered significant by the participating states. This approach yielded a robust set of indicators for assessing and comparing SQG performance across states. However, this approach does raise the general question of what is implied when some regulatory requirements are not included in an index - how important are these requirements for ongoing measurement of performance? In other cases, such as the Ontario Compliance Index, all requirements were included as indicators.

- **Keeping indicators up-to-date.** The relevance of indicators can change over time, which can make it necessary to update the set of indicators included in an index and/or revise how they are scored. For example, the Academic Performance Index is explicitly designed to address changing academic testing protocols over time. It includes two different measures of performance that are calculated each year: the base index, which reflects all indicators (i.e., both those from previous years and any new ones), and the growth index, which reflects only those indicators that were included in the previous year's score. As its name implies, the growth index is used to measure changes in performance since the previous year. In addition, as regulators adjust their indexes, they should consider how this affects their ability to compare results to indexes from previous years.

- **Impact on data collection and management.** Indicator selection can have significant implications for the effort required to collect and manage performance data. Incorporating new indicators of performance into an index may lead to additional effort both in training of inspectors and in updating information management systems to include new data fields. In the case of the IRS Compliance Index, the agency uses the same staff and data systems for both auditing and compliance measurement. The difference in the goals of these two activities lead to conflicts; for auditing the focused is on the overall outcome (i.e., how much additional tax money is recovered), while for measurement it is important to have complete information on the specific sources of any underpayment or noncompliance. The data management system is set up to accommodate the auditing function and does not require entry of audit results with the degree of specificity desired for measurement purposes.

DATA COLLECTION

Our review of indexing approaches identified the following key issues to consider when designing systems for data collection:

- **Selection of facilities to include in data collection.** To the extent that data collection is accomplished through inspections, regulatory requirements may determine which facilities are included (e.g., in the case of an industry composed primarily of major sources). For industries with large numbers of facilities, a statistical sampling approach may be desirable. Both ERP and the Common Measures Project use this approach, which allows for extrapolating the results to the broader population of facilities, in measuring performance in specific sectors.
- **Streamline data collection process.** Making on-site inspections as easy and objective as possible will improve the data reporting process. Standardizing inspection practices by using tools such as hand-held electronic data entry devices can speed up the inspection process and increase the accuracy of data analysis.[20,21] Connecticut tried to use this approach when collecting data on SQG performance, but experienced hardware compatibility issues. Another way to streamline data collection is to allow self-reporting of performance. The Massachusetts Capital Assets Management Deferred Maintenance Process relies on building managers to enter their facility data directly into a management information system. Similarly, facility managers are responsible for submitting data for four of the five attributes in Opra.
- **Level of subjectivity in data collection process.** In designing the data collection protocol, consider the potential for subjectivity on the part of the inspector (or other staff collecting the data). For example, when Connecticut was collecting data on SQG performance for the Common Measures Project, it hired interns and instructed them to rate each indicator on a numeric scale where five represented full compliance and decreasing numbers down to one represented more severe degrees of noncompliance. Ultimately, the interns lacked the experience and knowledge to distinguish the level of compliance and the final measure was modified to reflect simply either compliance or noncompliance.

WEIGHTING OF INDICATORS

Indicator weights can have a profound affect on the results of an index and reflect underlying priorities and value judgments concerning environmental performance. Based on our research, transparency and stakeholder involvement are highlighted below as important elements in developing successful weighting schemes.

---

[20] Metzenbaum, Shelley, Watkins, Allison, and Adeyeye, Adenkike. "A Memo on Measurement for Environmental Managers: Recommendations and Reference Manual". Environmental Compliance Consortium. September 2007.

[21] Metzenbaum, Shelley. "Guest Perspectives: Can Environmental Agencies Make Better Use of Accident, Incident, and Inspection Information". Spring 2007 ECOStates.

- **Importance of transparency in the weighting scheme.** Given the significant impact that indicator weights can have on the results of an index, it is critical that the underlying policy and regulatory priorities be transparent; otherwise, the weighting scheme may obscure the meaning of the index score. Ontario MOE has experienced some difficulty in conveying the meaning of its compliance index, with some stakeholders incorrectly associating higher scores with higher risk. In fact, while the weights represent to some degree the significance of a violation, they do not consider the environmental risk that a specific case of non-compliance poses. MOE is considering revisions to the current weighting scheme, which consists of only three different weights based on the significance of violations. The Academic Performance Index is characterized by a simple weighting scheme that clearly values good performance. It also includes a scaling system that translates numeric scores into a five categories, ranging from very poor to excellent. The Community Fluoride Compliance Index provides yet another example of transparency. The approach to calculating this index explicitly differentiates between non-reporting and noncompliance with the fluoride concentration standard, reflecting the greater significance of the latter.

  Another aspect of transparency involves the aggregation of indicators into a single value. While such a value can be useful for comparing and ranking performance among groups or facilities, it also can be important to preserve individual indicator scores to allow for more focused assessments of specific aspects of performance. Opra is designed to measure five facility attributes and provides individual scores for each of them. The hierarchy of indicators and weights in the EPI developed by Yale and Columbia (shown previously in Exhibit 2-7) allows users to consider results at multiple levels, including values for the single EPI score, two objectives, six policy categories, or 25 indictors.

- **Role of stakeholder outreach in developing weighting scheme.** Several of the case study indexes illustrate the important contribution of stakeholder engagement in developing weights for indicators. The first version of Opra developed by the EA, was subject to some criticism from facilities that considered the results to be too subjective and frequently challenged their scores. When the EA undertook an effort to update Opra, the agency met with facility representatives, sought public comment, and pilot tested changes being considered to help refine this tool by including more quantitative scoring and weighting of indicators and developing additional attributes to better reflect risk. In developing their EPI that scores environmental conditions in many different countries, researchers from Yale and Columbia held workshops, meetings and otherwise consulted heavily with experts in environmental science and policy. This level of engagement helped to ensure selected indicators and weighting scheme were appropriate across a diverse group of countries.

EXHIBIT 3-1: SUMMARY KEY LESSONS LEARNED

INDEX NAME/ SOURCE	WHAT IS MEASURED	KEY LESSONS FOR INDEXING
Academic Performance Index California School System	Academic performance	• Example of how weights can be applied to award "good" behavior, or theoretically to "punish" bad behavior; this allows the schools to be held accountable for their results. • Indicates importance of ensuring that results are comparable from year to year – provides for generation of a base API and growth API value each year to allow for adjustments in testing protocols.
Common Measures Project MA DEP. NEWMOA, and U.S. EPA	Performance levels of Small Quantity Generators (SQGs) on a series of compliance and beyond compliance indicators	• Illustrates considerations in selecting meaningful indicators – both regulatory and beyond compliance. • Presents scores for sector-wide performance levels across indicators and indicator-specific performance levels. • Example of how using common indicators and measurement approach can facilitate cross jurisdictional performance comparisons. • Demonstrates challenges of identifying relationship between agency activities and facility performance levels.
Community Fluoridation Compliance Index Journal Article	Performance of water systems in meeting regulatory requirement for fluoridation levels	• Shows how index results can be compared across time periods (because the denominator of the index equation allows for flexibility in time scales for measurement). • Raises issue of how to weight non-reporting vs. noncompliance: should a non-report be scored 0 for a sample/indicator or should it be adjusted based on the most recent reported value? • Weighting scheme reflects higher priority to protect against under fluoridation vs. over fluoridation. • Score is assessed by comparison to a rating system based on the population size served by the water system.
Disclosure Compliance Index Journal Article	Compliance with financial disclosure requirements	• Use of percent disclosure for all disclosure rules. Index includes adjustment of denominator to disregard non-applicable rules so that all non-disclosed values are related to noncompliance (not to rules not being applicable) • Indicators are scored 1 for full disclosure ("compliance") and 0 for nondisclosure. Index is calculated as a percentage of each firm's compliance with all of the relevant requirements.
ECC Compliance Rate Template Environmental Compliance Consortium	Facility compliance	• Shows how simple, cost-effective data collection and simple reporting techniques can incorporate supplemental information to help put the compliance rate in context and ensure accurate interpretation. • Reports compliance rates but emphasizes the importance of providing additional data to help interpret them, such as how facilities were selected for inspection and the total number of violations.

INDEX NAME/ SOURCE	WHAT IS MEASURED	KEY LESSONS FOR INDEXING
Environmental Performance Index (Portuguese Military) Journal Article	Self-assessed environmental performance for Portuguese military facilities	• The authors scored indicators and then translated weighted scores into a 5-category scale (ranging from "very poor" to "excellent") that clearly communicates the overall level of performance. • 18 indicators selected from 48 potential indicators based on two sets of criteria: relevancy and feasibility. • Used single continuous scale of variation that allows for weighting of indicators but authors elected to use equal weights for this analysis.
Environmental Performance Index Yale and Columbia	Global environmental performance	• This approach consolidates a large amount of data (25 indicators for 149 countries) and pares it down into meaningful, policy-relevant categories. • Groups 25 indicators into 6 policy categories, then 2 objectives, and finally an overall index. This hierarchy allows users to evaluate performance at multiple levels. • Indicator selection process involved consultation with many experts.
Environmental Results Program (ERP) US EPA	Facility performance levels on selected indicators of regulatory compliance and best practices	• Highlights the power of using statistically based random sampling to extrapolate the performance of a sample to the entire universe of facilities, and provides an approach for analyzing changes in performance over time. • Illustrates the role of statistical sampling and indicator selection. Key measure is change in percent of facilities meeting a specific requirement/indicator. • Conduct inspections to measure baseline, then inspect again to measure changes
IRS Compliance Index Internal Revenue Service	Voluntary filing compliance Reporting compliance Payment compliance	• Reveals the power of using statistically based random sampling to make inferences about population performance (i.e., the ability to extrapolate in cases with large populations). • The current NRP updated a previous system – had to minimize data collection burden and get stakeholders on board • Challenge: getting NRP data collection to work in context of daily audits. There are conflicting priorities – for audits, priority is to recover largest amount of money for least effort; for measurement, priority is to have most comprehensive data on compliance.
Massachusetts Capital Assets Management Deferred Maintenance Process MA Department of Capital Asset Management	Building condition and priority for specific maintenance projects	• Use of tiered indicators with weights that reflect priority areas for maintenance. • Example of comprehensive data management system with capability for building mangers to directly enter data on their buildings.
Ontario Compliance Index Ontario Ministry of the Environment (MOE)	Relative compliance levels	• Uses simple weighting scheme (three possible weights – 1, 2, and 4) to indicate the level of significance of violations (but does not capture risk). There are plans to revisit the weighting scheme and to expand use of index within MOE. • Based data collection on existing forms and processes; allows for use of current management information system. Change is in how data are reported

INDEX NAME/ SOURCE	WHAT IS MEASURED	KEY LESSONS FOR INDEXING
Operator Pollution and Risk Appraisal (Opra) UK Environment Agency (EA)	Potential risk associated with facility operations	• Example of an index that reflects a risk-based approach. It includes five attributes (complexity, emissions and inputs, location, operator performance, and compliance) that are all scored on a banded scale from A to E, with A being the lowest risk. • Facilities submit the data for 4 out of 5 attributes. EA enters compliance data based on its compliance classification scheme. • This system compiles data from several sources into a scoring system that allows the agency to "fairly" recover the costs of monitoring and targeting each facility. The scores also provide a way for the agency to track and assess the performance of facilities. • Initial version of Opra based on descriptive information for only two attributes: a pollution appraisal and operator performance. In response to concerns that the scoring was too subjective, EA updated Opra to include more objective and specific indicator scoring and also added more attributes.
Operator Safety Index (OSI) DOI Mineral Management Service (MMS)	Facility compliance with safety regulations and accident severity	• Using consistent indicators across multiple facilities with varying operations creates a "normalized" scale with which compliance and safety risk con be compared across the sector. • Weighting the instances of noncompliance by the severity of the enforcement action allows regulators to see the significance of noncompliance at each facility and create a "fair" comparison criterion.

REFERENCES

Ali, Muhammad Jahangir. "Disclosure compliance with national accounting standards by listed companies in South Asia." Accounting and Business Research, Vol. 34, No. 3, 2004. pp. 183-199.

Brown, Robert E. and Mazur, Mark J. "IRS's Comprehensive Approach to Compliance Measurement." IRS, June 2003. http://apps2.irs.gov/pub/irs-soi/mazur.pdf

California Department of Education. "2008-09 Academic Performance Index Report: Information Guide." May 2009, and LRC 2006-07 Documentation of Performance Index. http://www.cde.ca.gov/ta/ac/ap/documents/infoguide08.pdf

Esty, Daniel C., M.A. Levy, C.H. Kim, A. de Sherbinin, T. Srebotnjak, and V. Mara. 2008. 2008 Environmental Performance Index. New Haven: Yale Center for Environmental Law and Policy. http://www.yale.edu/epi/files/2008EPI_Text.pdf

Government Accountability Office. "Environmental Indicators: Better Coordination is Needed to Develop Environmental Sets that Inform Decisions". GAO-05-52, November, 2004. http://www.gao.gov/new.items/d0552.pdf

Hammond, Allen et al. "Environmental Indicators: A Systematic Approach to Measuring and Reporting on Environmental Policy Performance in the Context of Sustainable Development." World Resources Institute, May 1995. http://pdf.wri.org/environmentalindicators_bw.pdf

International Network for Environmental Compliance and Enforcement (INECE). "Principles of Environmental Compliance and Enforcement Handbook". April, 2009. http://www.inece.org/principles/PrinciplesHandbook_23sept09.pdf

International Network Environmental Compliance and Enforcement (INECE). "Performance Measurement Guidance for Compliance and Enforcement Practitioners: Second Edition". Expert Working Group on Enforcement and Compliance Indicators. April, 2008. http://www.inece.org/indicators/guidance.pdf

IRS National Research Program. "Challenges Associated with Collection Compliance Data." IRS Research Conference, 2002. http://www.irs.gov/pub/irs-soi/cocompda.pdf

IRS Office of Research, Analysis, and Statistics. "National Research Program: Early Results & Future Effort." Presentation, June 15, 2005. http://www.taxadmin.org/FTA/meet/05am_pres/brown.pdf

Kuthy, Raymond A. et al. "Use of a Compliance Index for Community Fluoridation." *Public Health Reports*, Vol. 102, No. 4, July-August 1987, pp. 415-420.

Metzenbaum, Shelley, Watkins, Allison, and Adeyeye, Adenkike. "A Memo on Measurement for Environmental Managers: Recommendations and Reference Manual". Environmental Compliance Consortium. September 2007.

http://www.nfwf.org/AM/Template.cfm?Section=Home&CONTENTID=13406&TEMPLATE=/CM/ContentDisplay.cfm

Metzenbaum, Shelley. "Compliance and Deterrence Research Project: Measuring Compliance Assistance Outcomes: State of Science and Practice White Paper". Prepared for U.S. EPA Office of Enforcement and Compliance Assurance. December 6, 2007. http://www.epa.gov/compliance/resources/reports/compliance/research/mcao-whitepaper.pdf

Metzenbaum, Shelley. "Guest Perspectives: Can Environmental Agencies Make Better Use of Accident, Incident, and Inspection Information". Spring 2007 ECOStates. http://www.ecos.org/files/2781_file_Spring_2007_ECOStates.pdf?PHPSESSID=607ea8092aef036e809b6dceccf98385

Metzenbaum, Shelley. "Measure to Comply, Measure to Perform: A Government Performance White Paper". Cognos, Inc. February, 2006. http://www.mccormack.umb.edu/centers/cpm/documents/measure_to_comply.pdf

Narang, Rajeev, Ontario Ministry of the Environment. "Compliance Index – Environmental Compliance/Performance Measure." Presented at the National Environmental Partnership Summit, Baltimore, MD. May 20, 2008. www.p2pays.org/ref/46/45184.ppt

Organisation for Economic Co-Operation and Development (OECD). "OECD Environmental Indicators: Development, Measurement, and Use." Reference Paper. http://unpan1.un.org/intradoc/groups/public/documents/apcity/unpan015281.pdf

Organisation for Economic Co-Operation and Development (OECD) EAP Task Force. "Recommendations on Performance Measurement for Environmental Enforcement and Authorities of Eastern Europe, Caucasus, and Central Asia." Eighth Annual Meeting of the EECCA Regulatory Environmental Programme Implementation Network. June 12/13, 2006. http://www.oecd.org/dataoecd/12/5/38137583.pdf

Organisation for Economic Co-Operation and Development (OECD). "OECD Key Environmental Indicators, 2004". OECD Environment Directorate, 2004. http://www.oecd.org/dataoecd/32/20/31558547.pdf

Ontario Ministry of the Environment. Compliance Index(CI): Draft Report. June 3, 2008.

Perini, David, Commissioner of Division of Capital Asset Management. FY 10 Deferred Maintenance Process, March 23, 2009.

Ramos, Tomas B. and de Melo, Joao Joanaz. "Developing and Implementing an Environmental Performance Index for the Portuguese Military." Business Strategy and the Environment, Vol. 15, 2006. pp. 71-86.

Shewmake, Tiffin. "Calculating and Communicating Environmental Compliance Rates." ECOStates Journal, Spring 2003. pp. 23-27. http://www.ecos.org/files/694_file_Spring_2003_ECOStates.pdf

Shewmake, Tiffin. "Using Compliance Rates to Manage." ECOStates Journal, Fall 2004. pp. 17-22. http://www.ecos.org/files/1327_file_Fall_2004_ECOStates.pdf

Shimshack, Jay P. "Monitoring, Enforcement, & Environmental Compliance: Understanding Specific & General Deterrence: Comparative Analysis of Monitoring and Enforcement Impact Measures". Prepared for the U.S. EPA Office of Enforcement and Compliance. June 2009. http://www.epa.gov/compliance/resources/reports/compliance/research/meec-whitepaper-task6.pdf

Slitor, Doug. "Measuring Safety and Compliance in the U.S. Offshore Oil and Gas Industry." *Society of Petroleum Engineers.* Paper prepared for presentation at the SPE International Conference on Health, Safety, and the Environment in Oil and Gas Exploration and Production held in Stavanger, Norway, June 26-28, 2000, SPE 61155.

Steven DeGabriele, Susan Peck and Tara Acker. The States Common Measures Project. June 19, 2009. http://www.newmoa.org/hazardouswaste/measures/finalreport/CommonMeasuresProjectFinalReport.pdf

The Environmental Council of the States. "State Environmental Agency Contributions to Enforcement and Compliance". Report to Congress, April 2001. http://www.ecos.org/files/687_file_ECOS_20RTC_20f.pdf

UK Environment Agency. "Spotlight on business – 10 years of improving the environment." July 2008. http://publications.environment-agency.gov.uk/pdf/GEHO0708BOFX-E-E.pdf?lang=_e

U.S. EPA, National Center for Environmental Innovation. "An In-Depth Look at the United Kingdom Integrated Permitting System." July 2008. http://www.epa.gov/permits/IntPermittingRpt.pdf

APPENDIX A

COMPLIANCE INDEXING CASE STUDY FACT SHEETS

1. Academic Performance Index (API)
2. Common Measures
3. Compliance Index for Community Fluoridation
4. Disclosure Compliance in South Asia
5. ECC Compliance Rate Template
6. Environmental Performance Index: Portuguese Military -
7. Environmental Performance Index - Yale Center for Environmental Law & Policy and CIESIN
8. Environmental Results Program (ERP)
9. IRS Compliance Measurement
10. Massachusetts Capital Asset Management Deferred Maintenance Process
11. Ontario Compliance Index
12. UK Operational Risk Appraisal Scheme (Opra)
13. Operator's Safety Index (OSI)

FACT SHEET | Academic Performance Index (API)

SPONSORING ORGANIZATION: California Department of Education

WESBITE AND REFERENCES: http://www.cde.ca.gov/ta/ac/ap/

California Department of Education. "2008-09 Academic Performance Index Report: Information Guide." May 2009, and LRC 2006-07 Documentation of Performance Index. Available at: http://www.cde.ca.gov/api.

PROGRAM DESCRIPTION:

MOTIVATIONS

The California Department of Education provides annual Academic Performance Index (API) reports as part of its Accountability Progress Reporting (APR) system. The APR system was developed as a way to report results of academic accountability programs for the state and federal accountability requirements in the California's Public Schools Accountability Act of 1999. The API measures a school's performance level based on the results of statewide standardized testing. It is intended to measure the academic growth of a school; the API from the current year is compared to scores from previous years to measure progress.

MEASUREMENT APPROACH:

WEIGHTING AND CONSTRUCTION

The index measures the academic performance and growth of schools on a variety of measures at the 3rd through 10th grade levels based on the number of students at each performance level. The API is calculated by converting a student's performance on statewide assessments across multiple content areas into performance levels on the API scale (i.e., untested, limited, basic, proficient, accelerated, and advanced). Each performance level is then assigned a weighted score. The percentage of students at each level is then multiplied by the respective weight, and the totals for each level are summed to get the school's overall performance index score.

For example, a school's overall performance index score would be calculated as follows:

PERFORMANCE LEVEL	% OF STUDENTS AT LEVEL	WEIGHT	SCORE
Untested	5	0.0	0.0
Limited	20	0.3	6.0
Basic	25	0.6	15.0
Proficient	35	1.0	35.0
Accelerated	5	1.1	5.5
Advanced	10	1.2	12.0
Performance Index Score			**73.5**

The percentage of students at each level is multiplied by the weight given to that level, which produces a score for that performance level. The overall performance index score is then calculated as the sum of the scores for each performance level. The maximum score in this scenario would be 120 (all students performing at the advanced level), and the minimum score would be zero (all students untested).

The index can be also be reported for individual schools by grade level, subject matter (i.e., content area), or performance level.

KEY ISSUES:

KEY STATISTICAL ISSUES, TOOLS, AND TECHNIQUES

One of the challenges of calculating this index from year to year is ensuring that results are comparable, or in other words, that the index is measuring the same thing every year. If, for example, the test significantly changed from one year to the next, adjustments need to be made to account for those changes. For instance, for the 2008-2009 school year, the California Achievement Test 6th Edition Survey was eliminated from the API, and the California Modified Assessment was added for grades three through five. Test weights or rules for inclusions/exclusions can also change from year to year. In order to account for these changes, the department produces two reports every year: the Base API and the Growth API. The Base API reflects all indicators (both those from previous years and new ones) and any methodological changes; this becomes the baseline against which the next year's API is compared. The Growth API for the current year is calculated to be consistent with the previous year's Base API and does not reflect new indicators or methodological changes. Base APIs are reported in the spring and Growth APIs in the fall.

USE AND RESULTS:

STRENGTHS, WEAKNESSES, AND BEST OPPORTUNITIES FOR USE

According to the latest API Reports Information guide, some of the key features of this index include the following:

- **Measures improvement.** The API is an effective way to measure the academic progress of a school from one year to the next. Each school has an annual target and all numerically significant subgroups at each school also have targets.
- **Subgroup accountability.** The API alerts schools to achievement gaps between traditionally higher- and lower-scoring student subgroups.
- **Cross-sectional picture of achievement.** The index does not track individual student progress across years but rather compares snapshots of school achievement results from one year to the next.
- **Ranking.** The API is used to rank schools by comparing scores statewide for schools with similar demographic characteristics.

FACT SHEET | Common Measures Project

SPONSORING ORGANIZATION: MA DEP, Northeast Waste Management Officials Association, U.S. Environmental Protection Agency

WESBITE AND REFERENCES: http://www.newmoa.org/hazardouswaste/measures/index.cfm

Steven DeGabriele, Susan Peck and Tara Acker. *The States Common Measures Project*. June 19, 2009.

PROGRAM DESCRIPTION: The Common Measures Project is designed to support state efforts to develop and use common measures of environmental performance for one or more business sectors/groups. It was funded by the 2005-2006 EPA State Innovations Grant Program (SIG).

Ten states participated in this effort - California, Colorado, Connecticut, Maine, Massachusetts, New Hampshire, New York, Rhode Island, Vermont and Washington – which applied the measurement methodology developed by Massachusetts for the Environmental Results Program (ERP) – see separate fact sheet for ERP for more details on this methodology.

States selected two industry groups for which to measure performance: small quantity generators of hazardous waste (SQGs) and auto body shops. They agreed on a common set of compliance and beyond compliance indicators for both groups and participated in data quality and statistical measurement training. States collected data from the SQG sector[22] using random sampling techniques to ensure that the data was representative for the entire universe of SQGs in each state. States then conducted statistical analysis to estimate how well each state's universe of facilities was performing on the indicators. This analysis enabled each state to assess whether performance levels were high enough on individual indicators and sets of indicators for SQGs in its jurisdiction, and how these facilities' performance compared to those in other states. A second level of analysis explored whether state activities influenced the measured SQG performance.

MOTIVATIONS

The purpose is to advance the use of valid statistical methods and measurement tools to enable comparison of performance changes across states resulting from the use of various environmental compliance assurance approaches. The stated project goals are to:

- Improve the ability of state environmental agencies to evaluate the performance of targeted business sectors including developing and

[22] As part of the project's grant commitment, participating states were required to complete analysis for at least one sector. Depending on resources, states could elect to work on more than one sector. States completed all work for the SQG sector. For the autobody sector, states selected indicators and developed a field checklist, but did not collect or analyze data. However, New York and Washington independently used the States Common Measures Project auto body indicators and field checklist to collect and analyze performance under a separate initiative.

implementing performance measures and using statistical approaches to analyze and report the results.

- Improve the ability of state environmental agencies to identify and adopt effective and efficient environmental performance improvement strategies based on those results.

NUMBER OF INDICATORS, WHAT THEY COVER, AND WHY THEY WERE SELECTED

States participating in the Common Measures Project selected a set of 12 indicator topics for SQGs, including eight regulatory and four beyond compliance indicators. In selecting these indicators, states evaluated a wide range of issues: including the different types of indicators to choose from (e.g., activity-based, outcome-based, regulatory, and beyond compliance). The indicators selected include:

REGULATORY INDICATORS

- Containers properly labeled.
- Containers closed.
- Containers in good condition.
- Accumulation quantity limits followed.
- Accumulation time limits followed.
- Manifests used.
- Hazardous waste streams identified.
- Emergency response information posted.

BEYOND COMPLIANCE INDICATORS

- Toxic use reduction implemented.
- Recycling projects undertaken.
- Water conservation implemented.
- Energy conservation/alternative energy implemented.

MEASUREMENT APPROACH:

DATA REQUIREMENTS TO COMPILE INDEX

The Common Measures Project is similar to ERP in the central role played by statistically based performance measurement. To accomplish this, each state collected data through on-site interviews at a sample of its SQGs.

CONSTRUCTION

Two types of scores were calculated for the Common Measures Project. The first is the SQG mean facility score, which can be based on "all indicators," "regulatory indicators," or "beyond compliance indicators." The facility score is the proportion of applicable indicators that the facility successfully achieved. It is measured on a scale of 0 – 10. A score of 10 indicates that the facility successfully achieved 100% of the indicators. A score of 0 indicates that the facility did not achieve any of the indicators. The mean facility score for a state is the average score for all facilities in that state.

The second score is the achievement rate. This is calculated for each performance indicator and represents the percentage of the facilities that achieved a specific indicator (i.e., complied with a regulatory requirement or implemented a beyond compliance practice).

KEY ISSUES:

KEY STATISTICAL ISSUES, TOOLS, AND TECHNIQUES

As with ERP, the random sampling approach is an important element of this project. It allows states to use information from the sampled SQGs to draw inferences about *all* facilities in the state, not just those facilities visited. Thus, by using statistics, states can use the observed data on SQG performance to estimate, within a certain range, the performance of all SQGs within each state.

When comparing results in SQG performance across states, the project report distinguishes between changes in performance that are statistically significant vs. those that are not. A statistically significant change in performance is one where one can be confident that an observed difference in SQG performance across one or more states occurred in the universe of facilities, not just in those included in the sample.

USE AND RESULTS:

STRENGTHS, WEAKNESSES, AND BEST OPPORTUNITIES FOR USE

Over a three year period, participating project states were able to use the same set of common measures to evaluate the annual environmental performance of a common group of facilities. The project also created a replicable template that can be used by other agencies to build the capacity to measure performance of a specific regulated group.

To provide insight into the influence of state policy activities on facility performance, participating states provided descriptions of the amount and nature of compliance and beyond compliance assistance provided, compliance inspection triggers and frequency, and enforcement tools and reporting requirements in place during the three years prior to the project. This information was compared to the performance results to identify if there were any oversight practices among the states that could be associated with higher performance rates. The results of this comparison suggested that there may be a relationship between compliance and beyond compliance assistance and SQG performance levels. While the data did not appear to support a relationship between the type of enforcement and measured performance levels, this methodology does not represent a definitive analysis and further investigation of this relationship is needed.

Looking forward, the Common Measures Project has paved the way for other state and federal agencies to undertake similar assessments. Current projects underway include a six-state initiative, in collaboration with EPA Region V (funded under the 2009 EPA State Innovations Grant Program), to use common indicators and statistical methods to assess region-wide performance of the auto body sector. In addition, EPA Region 1 and the EPA Office of Enforcement, Compliance and Assurance (OECA), are using statistical methods to compare performance of auto body shops in Massachusetts with those in a state where different policy approaches are being used.

FACT SHEET | Compliance Index for Community Fluoridation

SPONSORING ORGANIZATION: N/A – Academic Research

WEBSITE AND REFERENCES: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1477864/pdf/pubhealthrep00176-0069.pdf

Kuthy, Raymond A. et al. "Use of a Compliance Index for Community Fluoridation." *Public Health Reports*, Vol. 102, No. 4, July-August 1987. pp. 415-420.

PROGRAM DESCRIPTION: The authors of this article developed a Compliance Index for Community Fluoridation (CFCI), and then tested the index's usefulness in performing fluoridation surveillance at the state and local levels and to assess how much monthly fluoride concentrations vary.

MEASUREMENT APPROACH:

DATA REQUIREMENTS TO COMPILE INDEX

Fifty water systems in Illinois and 50 water systems in Ohio were randomly selected from all fluoridation systems in these states. Ohio and Illinois were selected because they both have fluoridation statutes, legal ranges for compliance, and readily available data. The data collected for each of these systems was:

- Monthly fluoridation levels;
- Population served by the water system; and
- The water plant operator's certification level.

EQUATIONS USED

The CFCI score for each community is based on the frequency of sampling per unit of time and degree of achievement of optimal concentration for each water system. Essentially, the CFCI calculates the average percentage of optimal fluoridation concentration over the number of samples that should have been submitted (e.g., monthly, weekly). It is calculated by taking the sum of the sample fluoridation concentrations divided by the optimal fluoridation concentrations. This sum is then divided by the number of samples that should have been submitted.[23]

$$CFCI = \frac{\sum\left(sample\_fluoridation\_concentration \big/ optimal\_fluoridation\_concentration \right)}{\#\_samples}$$

[23] Note that this equation represents the calculation for a sample in which the fluoride concentration is less than the optimal level. In the case where the sample fluoride concentration is greater than the optimal level, the numerator is flipped so that the optimal concentration is divided by the sample concentration.

In other words, the CFCI measures, on average, how close the water system is to achieving optimal fluoridation concentrations. The minimum score would be a 0.0, indicating that the sample fluoridation concentrations are zero. The maximum score would be a 1.0, indicating that the sample fluoridation concentration is at the optimal concentration level in every sample.[24] This score is then applied to a rating scale of system performance. The authors created this scale using the population served as the basis to establish the minimum acceptable score a water system can achieve to avoid administrative intervention. For example, a system serving 5,001 to 20,000 people has to achieve a CFCI score of 0.75 to avoid administrative intervention.

KEY ISSUES:

KEY STATISTICAL ISSUES, TOOLS, AND TECHNIQUES

The authors present three potential issues with the index:

1. Should equal weight be given to both the appropriate frequency of submitting samples and achieving the optimal fluoride levels? If a sample should have been received and was not, the system receives a score of zero for that sample, even if they were in fact in compliance.
2. The CFCI penalizes systems more for having suboptimal concentrations than for exceeding optimal levels. While there may be some concern with exceeding optimal levels, the greater concern is a system with suboptimal levels.
3. The index doesn't account for variations that may be out of the control of the water system, for example lower natural concentrations of fluoride in different regions of the country.

USE AND RESULTS:

STRENGTHS, WEAKNESSES, AND BEST OPPORTUNITIES FOR USE

The authors propose several benefits of this index:

- **Indicator of compliance.** It provides a ledger of compliance with state prescribed fluoridation concentration levels.
- **Flexible reporting.** The index can be reported monthly, quarterly, or annually. It can also be reported at the local, state, or national level.
- **Quick response.** Regulatory agencies can use it to quickly correct systems that stray from acceptable fluoridation concentrations.
- **Data requirements are flexible.** Regulators can change the optimal concentration levels and the number of months in the index for their specific needs.

[24] The authors also present a weighted average approach, in which the above calculations are modified so that a missing sample would be adjusted for using the previous month's sample.

FACT SHEET | Disclosure Compliance in South Asia

SPONSORING ORGANIZATION: N/A – Academic Research

REFERENCES: Ali, Muhammad Jahangir. "Disclosure compliance with national accounting standards by listed companies in South Asia." *Accounting and Business Research*, Vol. 34, No. 3, 2004. pp. 183-199.

PROGRAM DESCRIPTION: The authors of this article developed a total compliance index (TCI) that measures the level of compliance with disclosure requirements mandated by 14 national accounting standards for a large sample of companies within Bangladesh, India and Pakistan.

MOTIVATIONS

Level of disclosure of adequate and reliable information by companies in South Asia lags behind developed western capital markets and regulatory bodies are less effective in enforcing the existing accounting regulations. In addition, a lack of transparency and acceptance with internationally recognized reporting standards makes overseas investors hesitant to invest. The primary focus of this study was to determine the level of compliance with national accounting standards for a sample of 566 companies and to examine the corporate characteristics that are associated with compliance.

MEASUREMENT APPROACH:

WEIGHTING AND CONSTRUCTION

The index is compiled using the 14 common national accounting standards that have all been adopted by India, Pakistan and Bangladesh. These 14 standards were used to develop a checklist of 131 disclosure compliance requirements. Each company was given a score of zero for each requirement it does not disclose and a score of one for each requirement that it does. The authors did not apply any weights to these individual requirement scores. The total compliance index (TCI) is then calculated as the ratio of the total disclosure score (sum of each individual requirement score) to the expected number of items required to be disclosed. The TCI is therefore a percentage of each firm's compliance with all relevant requirements. A minimum score would be zero, indicating that the firm is out of compliance with all relevant requirements. The maximum score would be 100 percent, meaning that the firm is in compliance with all relevant requirements.

KEY ISSUES:

KEY STATISTICAL ISSUES, TOOLS, AND TECHNIQUES

One of the issues with calculating this index is the distinction between a non-disclosure that is a result of non-compliance versus one that is a result of non-applicability. The authors resolved this issue by making the score relative only to the requirements for which the firm is responsible. If this approach had not been applied, several scores would have been artificially low, as non-applicable items would be treated as instances of non-disclosure.

FACT SHEET | ECC COMPLIANCE RATE TEMPLATE

SPONSORING ORGANIZATION: Environmental Compliance Consortium

REFERENCES: Shewmake, Tiffin. "Calculating and Communicating Environmental Compliance Rates." *ECOStates Journal*, Spring 2003. pp. 23-27.

Shewmake, Tiffin. "Using Compliance Rates to Manage." *ECOStates Journal*, Fall 2004. pp. 17-22.

PROGRAM DESCRIPTION:

MOTIVATIONS:

The Environmental Compliance Consortium (ECC) has developed an Excel-based template for calculating compliance rates. This effort is meant to address the fact that agencies cannot afford to inspect every regulated facility on a regular basis and must therefore select a subset to inspect. A state may use different methods for selecting the subset (i.e. random sampling, targeted or complaint-driven inspections) which are likely to influence the resulting compliance rate. ECC's Compliance Rate Template calculates a compliance rate and also provides additional information critical to the rate's interpretation (e.g., whether the inspections were pre-announced, reasons for monitoring compliance, etc.), providing consistency across states, programs, and industry sectors.

MEASUREMENT APPROACH:

DATA REQUIREMENTS

The Compliance Rate Template presents a compliance rate along with additional information necessary to interpret the rate:

- Number of Compliance Monitoring Actions (CMAs)
- Reason for CMAs
- Methodology for choosing facilities for compliance monitoring
- Number of facilities where a CMA was conducted
- Number of facilities in the sector, facility type, or program focus
- Number of facilities with violations
- Total number of violations
- Number of facilities with significant violations (based on EPA definitions)
- Total number of significant violations

ECC's template provides six common reasons why states conduct CMAs. These reasons include: suspected problem; correction check-up; complaint-driven; regularly scheduled;

geographic initiative; and sensitive ecosystem/critical environment area. ECC has also identified common methods for choosing which facilities to monitor: monitoring the entire universe of regulated facilities; monitoring a statistically valid, randomly sampled subset of the universe; and monitoring a subset of the entire universe chosen in another manner.

KEY ISSUES:

KEY STATISTICAL ISSUES, TOOLS AND TECHNIQUES

The Compliance Rate Template does not require the use of complex or costly statistical methods or models. Compliance rates are presented along with the information on how they were obtained but the rates themselves are not altered or adjusted by that information.

USE AND RESULTS:

PAST AND CURRENT USE

Several states, such as New jersey and Delaware, have used the Compliance Rate Template and have found it useful for evaluating program effectiveness and allocating resource.

Other states have calculated compliance rates not necessarily using the template but adopting some of its key characteristics. These states have used the template's underlying concepts to report compliance in alternative ways from a straight compliance rate. For example, Colorado considers a facility that has had one or more violations during the year to be out of compliance. In addition to capturing such compliance rates, however, Colorado tracked the number of violations per year. It found that after conducting inspections more frequently, the compliance rate remained approximately the same, but the number of violations per year decreased significantly.

STRENGTHS, WEAKNESSES; AND BEST OPPORTUNITIES FOR USE

ECC's Compliance Rate Template allows states to transparently calculate compliance rates and communicate relevant interpretive information. However, the template has limitations, such as:

- It does not reflect variations in the number of compliance obligations across facilities (i.e., there is no difference between a facility that failed to meet five out of five requirements and a facility that failed to meet 25 out of 200 requirements);
- It does not communicate the duration of a facility's noncompliance; and
- It does not reflect the percentage of noncompliant facilities at any given point in time.

The ECC considers this effort a work in progress, and expects that approaches to calculating and communicating compliance rates will continue to evolve.

FACT SHEET | Environmental Performance Index - Portuguese Military

SPONSORING ORGANIZATION: N/A (Academic Research)

CONTACT: None

WEBSITE/ REFERENCES: Ramos, Tomas B. and de Melo, Joao Joanaz. "Developing and Implementing an Environmental Performance Index for the Portuguese Military." *Business Strategy and the Environment*, Vol. 15, 2006. pp. 71-86.

PROGRAM DESCRIPTION: The authors of this article had two goals:

1. To assess the Portuguese military sector's environmental performance through an index; and
2. To compare the military's self-assessment against the environmental profile created by the index to determine how well they matched.

MOTIVATIONS

The Military Environmental Performance Evaluation (MEPE) index was proposed by university researchers to the Portuguese Ministry of Defense as a tool to help decision-makers understand the sector's environmental impacts and performance.

One of the motivations for developing this index is the importance of environmental performance measurement and communication with the public, the military's primary stakeholder. This is especially relevant in situations where military activities affect the environment beyond the borders of military facilities.

NUMBER OF INDICATORS, WHAT THEY COVER, AND WHY THEY WERE SELECTED

The index measures the extent to which environmental practices have been adopted by the military. A survey with 48 questions was administered, of which 18 were selected as indicators to be aggregated into the index based on two sets of criteria: relevancy and feasibility. The relevancy criteria relates to an indicator's:

- Technical and scientific importance;
- Synthesis capability (i.e., ability to compare the indicator across sectors);
- Usefulness for communicating and reporting;
- Reflection of major environmental practice issues in defense organizations; and
- Importance to the environmental head in the defense sector.

The feasibility criteria relates to an indicator's:

- Robustness;
- Availability to future assessments; and
- Non-confidentiality.

MEASUREMENT APPROACH:

DATA REQUIREMENTS TO COMPILE INDEX

The index was developed using data from a national questionnaire, which was administered to evaluate the environmental practices adopted by 133 military units in the Army, Air Force and Navy. The questionnaire was based on self-assessment and contained 48 questions about environmental performance, including topics such as:

- Appointment of a person responsible for the environment,
- Environmental training for personnel,
- Knowledge and implementation of environmental management systems,
- Environmental programs,
- Environmental cooperation with stakeholders, and
- Environmental standards for suppliers.

EQUATIONS USED

Once the indicators were chosen, the raw data was transformed into a single continuous scale of variation, from zero (worst performance) to one (best performance). The index was calculated using the following equation:

$$MEPE = \sum_{j=1}^{m} \frac{\left[\frac{1}{\sum w} \sum_{i=1}^{n} X_i w_i \right]_j}{m}$$

Where,

X_i = the indicator of environmental practice *i* derived from the questionnaire, which is attributed a relative weight, *w*

n= the total number of *i* indicators

m = the total number of *j* units

WEIGHTING

The approach used to aggregate the indicators' scores does allow for different weights to be applied to different indicators to account for their relative importance. In this paper, however, the authors assigned identical weights to each of the 18 indicators. The index is then calculated using an aggregation process, with the end result being a value from zero to one.

The authors then established five categories to classify the range scores from the index (0.0 to 1.0), from very poor to excellent (See Exhibit 1).

EXHIBIT 1: CATEGORIES OF ENVIRONMENTAL PERFORMANCE

INDEX SCORE	ENVIRONMENTAL PERFORMANCE CATEGORY
0.0 – 0.20	Very Poor
0.21 – 0.40	Poor
0.41 – 0.60	Medium
0.61- 0.80	Good
0.81 – 1.0	Excellent

USE AND RESULTS:

RESULTS

Overall, the index showed that the Portuguese military sector's environmental performance is poor, with a score of 0.33. According to their scale, an index score of 0.33 means that the sector has a "poor" environmental performance.

The survey also asked a self-assessment question about the overall performance of the military with regard to environmental practices. The question asked individuals to rank performance from zero (very poor) to five (excellent). The answers to this scale were directly compared with the environmental performance categories based on the index scores to determine how well they match. The researchers found that the self-assessment results were overly optimistic when compared to the index results.

The results of the index are also presented by military branch, land area class, and personnel class. The personnel classes are broken out by population size. They find that units with smaller populations under-perform relative to those with larger populations.

FACT SHEET | ENVIRONMENTAL PERFORMANCE INDEX - 2008

SPONSORING ORGANIZATION: Yale Center for Environmental Law & Policy; Center for International Earth Science Information Network (CIESIN)

WEBSITE AND REFERENCES: http://epi.yale.edu

Esty, Daniel C., M.A. Levy, C.H. Kim, A. de Sherbinin, T. Srebotnjak, and V. Mara. 2008. 2008 Environmental Performance Index. New Haven: Yale Center for Environmental Law and Policy.

PROGRAM DESCRIPTION:

MOTIVATIONS

The Environmental Performance Index (EPI) is an analytical tool used to measure the environmental performance across individual nations. The EPI is meant to address data gaps that impede cross-country comparisons and policy analysis within each country. It provides data on actual outcome measures related to the environmental issues that many governments have prioritized.

STAKEHOLDER INVOVLEMENT

To develop the EPI and gather relevant data, the lead authors from Yale and CIESIN consulted with subject-area experts, statisticians, and policymakers from around the world.

NUMBER OF INDICATORS, WHAT THEY COVER, AND WHY THEY WERE SELECTED

The EPI is built using 25 individual indicators. Some of these indicators are direct measures of environmental performance and some are proxies. The 25 indicators cover environmental challenges identified by policy and scientific experts. They were chosen to represent a broad range of specific environmental challenges. To select the indicators, the authors of the EPI relied on a review of the environmental science literature, the advice of scientific and policy experts, and assessments conducted by international agencies and non-governmental organizations. The authors chose indicators using four selection criteria: relevance, performance orientation (meaning the indicator tracks outcome measures), transparency, and data quality. Examples of indicators include: Drinking Water, Indoor Air Pollution, Critical Habitat Protections, and Industrial Carbon Intensity.

MEASUREMENT APPROACH:

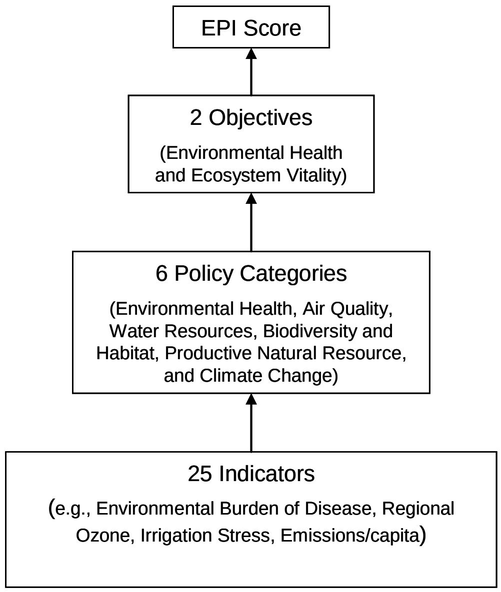
DATA REQUIREMENTS TO COMPILE INDEX

The EPI reports environmental performance scores for 149 countries. The authors gathered the most reliable data available, drawn primarily from international, academic, and research institutions.

WEIGHTING AND CONSTRUCTION

Each indicator is measured using a “proximity-to-target value.” The EPI authors identified specific targets for each indicator, typically based on international treaties and agreements or other widely accepted standards. The indicator is assigned a value on the scale of zero-100, with 100 meeting the target level and zero being the worst observed value. These indicators are then weighted and aggregated to produce scores for six policy categories: Environmental Health, Air Quality, Water Resources, Biodiversity and Habitat, Productive Natural Resource, and Climate Change. These six policy categories are weighted and aggregated again to produce scores for two main objectives: Environmental Health and Ecosystem Vitality. Finally, the overall EPI is calculated as the arithmetic mean of the Environmental Health and Ecosystem Vitality objective scores. See Figure 1 for an illustration of how the EPI is constructed.

Figure 1. Construction of the EPI



Groupings and weights were identified using a principal component analysis (PCA).[25] If a PCA basis for weighting the indicators was not available, the authors used equal weights with some refinements based on expert guidance. The equal division of the Environmental Health and Ecosystem Vitality categories represents a policy judgment on the part of the authors. Other indicator weights were also based on common policy priorities, scientific consensus, and reliability.

KEY ISSUES:

KEY STATISTICAL ISSUES, TOOLS AND TECHNIQUES

The authors examined the distribution of each indicator to identify extreme outliers, then applied winsorization – a statistical technique in which the authors trimmed outliers at the 95th percentile of the distribution.

In cases where a country’s performance was better than the identified target, the country was assigned a score of 100 – countries were not “rewarded” for environmental performance beyond the standards used to set indicator targets.

[25] PCA weights are statistically derived to reflect the importance of a given indicator relative to others with respect to the principal components.

DATA COLLECTION

The EPI remains limited by available data. Almost 90 countries are not included in the 2008 EPI because they lack the quality and quantity of data required to calculate an EPI score. Data gaps prevent the inclusion of measures for such relevant issues as:

- Exposure to toxic chemicals and heavy metals
- Waste management (including both household and toxic waste)
- Nuclear safety
- Pesticide safety and chemical exposure
- Wetlands loss
- Agricultural soil quality and erosion

The current EPI does not include these indicators, but the authors hope that they will be incorporated into the index in the future.

USE AND RESULTS:

HISTORY AND CURRENT USE

The EPI was compiled in a collaborative effort between the Yale Center for Environmental Law and Policy, CIESIN, the World Economic Forum, the Joint Research Centre of the European Commission, and various experts. The goal of the index is to support the use of accurate data and analysis in environmental policymaking. The EPI has been reported on and used by mainstream media outlets, such as National Geographic and the New York Times.

STRENGTHS, WEAKNESSES; AND BEST OPPORTUNITIES FOR USE

The EPI compiles environmental data that can be used to address national policy objectives. It can be used to identify problems and prioritize environmental issues, indicate where environmental policy is effective or ineffective, obtain an objective indication of the level of environmental threats, provide a baseline to compare one country's performance to another, and track environmental performance trends. The EPI produces overall results but also results by political, geographic, and economic peer groups. Additionally, the EPI provides results by clusters of countries having similar results for the individual indicator scores.

FACT SHEET | Environmental Results Program (ERP)

SPONSORING ORGANIZATION: The following states have implemented a full ERP, with the support of the U.S. Environmental Protection Agency, at this time: Delware, Florida, Illinois, Indiana, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Nevada, New York, Rhode Island, Vermont, Virginia, Washington, and Wisconsin.

WESBITE: http://www.epa.gov/erp/

PROGRAM DESCRIPTION: ERP is an innovative approach to improving the environmental performance for sectors or groups of regulated entities characterized by large numbers of small, relatively similar facilities. ERP combines plain language compliance assistance that promotes pollution prevention; facility self-assessment and self-certification; agency inspections; and statistically based performance measurement. Where necessary, regulators also conduct a comprehensive facility inventory and targeted enforcement actions.

The steps involved in developing and implementing a typical ERP can vary, depending on the specific needs of the regulator; therefore, ERPs may include all or some of the steps outlined below. In addition, states vary regarding whether they make these steps mandatory or voluntary, depending again on the needs of the regulator. Generally, for a full ERP, states:

1. **Identify facilities in the target sector.** In this step, states identify the names and locations of all relevant facilities in the entire state (or focused area), to the best of their ability.
2. **Conduct inspections at a random sample of facilities at the outset of the program (i.e., establish baseline).** Based on the number of facilities identified in the target sector, the state then determines the number of facilities that must be inspected to obtain a statistically valid random sample. Inspections are conducted at this random sample of facilities, during which inspectors complete checklists that describe the extent to which each facility's performance adheres to both regulatory and beyond compliance measures.
3. **Offer compliance assistance to all facilities.** There are several forms of compliance assistance that states can offer. Common examples include workbooks and workshops. These materials generally describe the regulatory requirements and beyond compliance practices for the sector and serve as a basis for self-certification.
4. **Encourage (or in some cases require) facilities to conduct a self-assessment and submit self-certification forms.** Facilities use the materials provided during the compliance assistance (e.g., workbooks) to assess the extent to which they are in compliance with their regulatory requirements, using the self-certification forms.

5. **Conduct a second round of random inspections**. Once the compliance assistance has been administered and the self-certification forms have been completed, the state will conduct a second round of random inspections to again determine the extent to which each facility's performance adheres to both regulatory and beyond compliance measures.

6. **Compare results of baseline inspections to the second round inspection results**. The performance results of the baseline and second round of results are then compared to determine the performance changes of the sector as a whole.[26]

7. **Utilize performance data to inform and improve future compliance assistance**. States can use the results of this analysis to target resources toward sector issues or behaviors that are of particular concern. For example, if a certain set of indicators showed a decline or very little improvement in performance, more resources may be needed in that area.

MOTIVATIONS

These components are intended to work together to improve compliance and reduce the environmental impacts of the target sector while deploying government resources strategically and efficiently. ERP is an integrated approach that often addresses multiple environmental media, and combines efforts involving compliance assistance and measurement. Facilities receive a comprehensive package of information from the state such as a workbook describing regulatory requirements, best practice suggestions, and self-certification forms.

NUMBER OF INDICATORS, WHAT THEY COVER, AND WHY THEY WERE SELECTED

At the outset of an ERP, state regulators must decide which indicators they want to include in their inspections and self-certification forms. Often, this set of indicators includes both regulatory requirements and beyond compliance best practices. Then a detailed checklist of these indicators is created for on-site inspections.

MEASUREMENT APPROACH:

DATA REQUIREMENTS TO COMPILE INDEX

A key component of the ERP design is statistically based performance measurement. To achieve this, regulators conduct inspections at a random sample of facilities both before and after any compliance and/or outreach efforts. During site visits, inspectors assess the performance of the facility by filling out the detailed checklists that indicate whether or not the facility is following the applicable compliance and pollution prevention practices selected.

CONSTRUCTION OF SAMPLES

The state then calculates the percentage of shops in the sample following each practice at the baseline; this is called the observed proportion of shops. Next the state conducts its outreach component (e.g., by holding compliance assistance workshops). After the outreach phase, facilities are given time to implement changes. Then the state measures performance again by surveying a sample of facilities using the same checklists, allowing

[26] Conducting random inspections of a sample of facilities allows states to make inferences about the performance of the sector as a whole, with statistically significant findings, rather than having to inspect every facility in the sector.

it to calculate the observed proportion of shops following each practice. The state can then compare the two sets of inspection results to determine whether the percentage of shops following each practice changed relative to the baseline.

KEY ISSUES:

KEY STATISTICAL ISSUES, TOOLS, AND TECHNIQUES

Importantly, the random sampling approach allows states to use information from the sampled facilities to draw inferences about *all* facilities in the universe, not just those facilities visited by inspectors. Thus, by using statistics, states can use the observed proportion of shops in the sample following each practice to estimate, within a certain range, the proportion of shops in the *universe* following that practice. Moreover, the state can compare estimates of the universe's baseline performance to estimates of its post-certification performance to assess whether or not the overall percentage of facilities in the entire population following specific practices has changed and if so, by how much.

When reporting results from ERPs, states distinguish between changes in performance that are statistically significant versus those that are not. A statistically significant change in performance is one where states can be confident that a change in performance occurred in the universe of shops, not just in the sample of facilities visited. In other words, if a statistically significant change occurred, a state can be confident that the percentage of *all* shops in the universe following a certain practice is different at baseline and post-certification.

Saying that a change is statistically significant does not, however, indicate the degree of change that occurred. Rather, to understand how much performance changed, regulators need to consider the magnitude of difference between the baseline and post-certification inspections.

USE AND RESULTS:

STRENGTHS, WEAKNESSES, AND BEST OPPORTUNITIES FOR USE

ERP results can be compiled in a number of ways: for example, by environmental medium, regulatory requirements, geographic area, or best practice category. Exhibit 1 below provides an example of how ERP performance changes can be aggregated and reported, by environmental medium.

EXHIBIT 1: EXAMPLE PERFORMANCE RESULTS

ENVIRONMENTAL MEDIUM	# OF INDICATORS	# IMPROVING (# SIGNIFICANT)	# WORSENING (# SIGNIFICANT)	# NO CHANGE
Air Emissions	14	9 (6)	3 (0)	2
Waste Management	19	16 (9)	2 (0)	1
Water Discharge	9	5 (4)	4 (0)	0
Worker Health and Safety[(b)]	5	3 (3)	2 (0)	1
Overall	**47**	**33 (22)**	**11 (0)**	**4**

FACT SHEET | IRS Compliance Measurement

SPONSORING ORGANIZATION: U.S. Internal Revenue Service

CONTACT: Bob Brown, Office of Research, Analysis and Statistics

WESBITE AND REFERENCES: http://www.irs.gov/privacy/article/0,,id=139179,00.html

IRS National Research Program. "Challenges Associated with Collection Compliance Data." IRS Research Conference, 2002.

Brown, Robert E. and Mazur, Mark J. "IRS's Comprehensive Approach to Compliance Measurement." IRS, June 2003.

IRS Office of Research, Analysis, and Statistics. "National Research Program: Early Results & Future Effort." Presentation, June 15, 2003.

PROGRAM DESCRIPTION:

MOTIVATIONS

In 1998 the IRS created the National Research Program, a comprehensive effort to measure and report taxpayer compliance. The program was created as a response to the Reform and Restructuring Act of 1998 which mandated that the IRS deliver services to taxpayers at a new and much higher level of performance. As part of this effort, the IRS saw a need for reliable, regular, and up-to-date approaches to measure overall compliance to allow the IRS to rely on compliance measures as strategic performance indicators. The NRP was seen as a way to increase public confidence in the fairness of the tax system by helping the IRS to identify where compliance problems occur and focus its resources accordingly.

NUMBER OF INDICATORS, WHAT THEY COVER, AND WHY THEY WERE SELECTED

The program does not use an index but rather a multi-prong measurement system that covers three types of mutually exclusive measures of taxpayer compliance that together, provide a comprehensive look at overall taxpayer compliance:

1. **Reporting compliance**: accuracy of tax timely reported,
2. **Filing compliance**: the percent of returns timely filed, and
3. **Payment compliance**: the percent of reported tax liability timely paid.

These compliance measures are strategic measures because they support the IRS's strategic goal of improving compliance. Operational measures, in contrast, focus on enforcement activities associated with compelling noncompliant taxpayers to meet their tax obligations.

MEASUREMENT APPROACH:

NRP has developed two measures of filing compliance: the filing compliance rate and the nonfiling tax gap. The filing compliance rate captures the frequency of voluntary filing compliance and the nonfiling tax gap captures the dollar amount of filing noncompliance.

EQUATIONS USED

Filing compliance rate: the percentage of the taxpayer population with a filing requirement that filed timely.

$$\textit{Filing rate} = \frac{\text{Number of required returns filed timely}}{\text{Total number of returns required to be filed}}\ X\ 100$$

Nonfiling tax gap: the dollar amount of unpaid taxes due from delinquent and non-filed returns.

NRP has also developed two measures of payment compliance: the voluntary payment compliance rate (VPCR) and the cumulative payment compliance rate (CPCR).

Voluntary payment compliance rate (VPCR): the percentage of the total tax paid timely on timely filed returns relative to the total tax reported on timely filed returns.

$$VPCR = \frac{\text{Tax paid timely on timely filed returns}}{\text{Tax reported on timely filed retuns}}\ X\ 100$$

Cumulative Payment Compliance Rate (CPCR): percentage of tax paid on timely filed returns to date relative to the total tax reported on timely filed returns.

$$CPCR = \frac{\text{Tax paid to date on timely filed returns}}{\text{Tax reported on timely filed returns}}\ X\ 100$$

Finally, the NRP has developed three measures of reporting compliance: the voluntary reporting rate (VRR), the underreporting gap, and the net misreporting percentage (NMP).

Voluntary Reporting Rate (VRR): percentage of tax that taxpayers accurately reported on timely filed returns.

$$VRR = \frac{\text{Total tax reported on timely filed returns}}{\text{Total tax reported} + \text{estimate of tax misreported}}\ X\ 100$$

Underreporting gap: amount of tax that is not voluntarily reported on timely filed returns for a given tax year. This value represents the extent to which timely filers understand their true tax liability, and the net of their overstatements.

Net Misreporting Percentage (NMP): the amount misreported on each line divided by the amount that should have been reported.

$$NMP = \frac{\text{Sum of the amounts misreported on the income or offset item}}{\text{Sum of the absolute values of the amounts that should have been reported on the income or offset item}} X\ 100$$

KEY ISSUES:

DATA COLLECTION

The NRP reporting studies rely on direct examinations of randomly selected returns. The first study relied on data from 46,000 tax year 2001 returns.[27]

Two of the major improvements the program has made to its compliance measurement data collection are: better use of databases and electronic information and partnering with U.S. Census and other external data to reduce taxpayer reporting burden. The former allows auditors to build case files that allow the agency to track information from year to year and complete the audits more efficiently. The latter gives them the ability to confirm several pieces of information to minimize the evidentiary burden on the taxpayer, e.g., confirm marriages without having to see actual marriage licenses.

The program has had to face several data collection issues in the development of its compliance rates, including:

- Availability of a data source for tax entities other than individuals (e.g., small businesses that do not request an EIN).
- Time required to develop data (release of data occurs 18 months after end of tax year).
- Privacy and sensitivity issues such as IRS access to Census data.

KEY STATISTICAL ISSUES, TOOLS, AND TECHNIQUES

The program has also had to face several statistical issues in the development of its compliance rates, including:

- Statistical reliability of the data and the data-extract programs.
- Definitions of what is (or is not) a payment, and what is (or is not) a late payment.
- Research vs. reality, i.e., the conflict between demands on the IRS research communities and the political and practical constraints on collecting compliance measurement data.
- Internal communication.
- External communication (stakeholder involvement).

---

[27] The first time this analysis was completed, it took three years to finish for one tax year. Moving forward, the NRP analyzes one tax year at a time and now can produce results in less time.

- Research audits differ from day to day audits. Day to day auditors are interested in going in, receiving the most payments they can, and getting out. Auditors may not collect data for every line of a tax return if the big ticket items have been cleared up. From a research standpoint, auditors should collect every piece of data possible to obtain a complete data set from each tax return.
- Small adjustments mean large overall impact (each tax return in the sample represents 10,000 taxpayers).
- IT infrastructure. The primary purpose of the current reporting system is to end up with a value the taxpayer owes; the system allows for auditor discretion, as long as the end result is accurate. For example, if small business income is supposed to be entered in three places throughout the tax return, but only entering it in one still allows the system to report the correct income earned, auditors do not need to enter the data in all three places. From a research standpoint, they would like every line item to be entered correctly. They would prefer that the system force auditors to enter every line item.

USE AND RESULTS

HISTORICAL AND CURRENT USE

Starting around 1960, the IRS used a system called the Taxpayer Compliance Measurement Program (TCMP) that primarily focused on individual tax returns. A representative sample of approximately 50,000 returns was randomly selected every three years to identify potential noncompliance. Based on these audits, the IRS would model how likely different segments of the population are to be noncompliant (e.g., by income level). In the 1970s and 1980s, the agency started focusing on tax gap analysis, the difference between taxes paid and taxes owed for all federal taxes and all taxpayers. The last TCMP study of voluntary filing compliance occurred in 1988, and the last study of voluntary payment compliance occurred in 1984.

The NRP expanded and re-designed the TCMP to include two new measures of compliance and to change the way that data was collected. The new guiding principles of the NRP were to:

- Minimize taxpayer burden related data collection;
- Involve the Business Operating Divisions as partners in the design and implementation of the program, as well as customers of the results; and
- Solicit external stakeholder (e.g., OMB, Congress, Treasury, GAO) ideas and support in the design of the program.

STRENGTHS, WEAKNESSES, AND BEST OPPORTUNITIES FOR USE

The program now produces annual studies that report the compliance rates described above. In addition to the annual reports, the NPR reports that its data allow the agency to:

- Conduct analyses of market segments (e.g., high income individuals);
- Prepare preliminary estimates of the tax gap;

- Improve the IRS' ability to detect noncompliance;
- Reduce the burden of unnecessary IRS contact with compliant taxpayers;
- Support strategic decision-making, program development and resource allocation of the IRS Operating Divisions;
- Provide the Commissioner and other senior IRS executives with an indication of the current state and recent trends of the voluntary compliance behaviors of U.S. taxpayers; and
- Provide benchmarks against which the IRS can evaluate the effectiveness of programs designed to improve taxpayer compliance with the tax code.

FACT SHEET | MASSACHUSETTS CAPITAL ASSET MANAGEMENT DEFERRED MAINTENANCE PROCESS

SPONSORING ORGANIZATION: MA Division of Capital Asset Management (DCAM)

CONTACT: Hope Davis, Director of Facilities Maintenance, DCAM

WEBSITE AND REFERENCES: Website for Executive Office for Administration and Finance, Facilities Management and Maintenance
http://www.mass.gov/?pageID=afsubtopic&L=3&L0=Home&L1=Property+Management+%26+Construction&L2=Facilities+Management+%26+Maintenance&sid=Eoaf

David Perini, Commissioner of Division of Capital Asset Management. *FY 10 Deferred Maintenance Process*, Memo, March 23, 2009.

PROGRAM SUMMARY: MA DCAM is responsible for administering the Massachusetts program for maintaining state-owned building facilities. DCAM established a database, the Capital Asset Management Information System (CAMIS) along with a process to assist in identifying maintenance and renovation needs and allocating scarce resources across approximately 5,000 state buildings, representing 64 million square feet. The State Legislature directed the agency to conduct a condition assessment of state-owned buildings in the late 1990s. Data on the quality of 12 different building systems (e.g., roof, exterior walls) were compiled into CAMIS. State agencies update data in CAMIS on an ongoing basis for their facilities.

MOTIVATIONS

The Patrick Administration wanted to centralize activities related to maintenance of facilities across all agencies. In response, DCAM developed a system for scoring annual deferred maintenance projects. This system is based on CAMIS and a set of prioritization criteria that serve to help allocate financial resources for building maintenance and also to demonstrate that these resources are being spent wisely.

STAKEHOLDER INVOLVEMENT

DCAM developed a straw proposal for criteria and weights and sought feedback from facility managers.

NUMBER OF INDICATORS, WHAT THEY COVER, AND WHY THEY WERE SELECTED

The scoring system consists of 11 quantitative indicators in three tiers. They cover a wide range of issues related to building deficiencies, including: the level of danger posed by them, urgency of work needed, and threat of penalty.

MEASUREMENT APPROACH:

DATA REQUIREMENTS TO COMPILE INDEX

The data for facilities initially came from the condition assessment completed in the early 2000s and compiled in CAMIS. Since then, facility managers have had the opportunity to update information on facility conditions. DCAM sends out the current information on an annual basis in CAMIS, to agencies managing facilities so they know what is in the database and can make any necessary updates. DCAM offers training programs on the use of CAMIS for staff from all agencies with facilities in the system.

WEIGHTING AND CONSTRUCTION

The annual deferred maintenance scoring is based on a set of 11 prioritization criteria (see Exhibit 1). These criteria are split into three tiers with weighting factors of 2.0, 1.5 and 1.0. The top score based on the rating scale and weights is 28. In the case of a tie, there are also a series of tier four criteria, including cost, that are considered when allocating funding.

KEY ISSUES:

KEY STATISTICAL ISSUES, TOOLS AND TECHNIQUES

Maintenance capital funding requests are evaluated based on the scores for proposed projects according to tiered prioritization criteria (see Exhibit 1).

USE AND RESULTS:

HISTORY OF USE

DCAM uses CAMIS and the prioritization criteria to evaluate funding requests This system requires annual data input by state agencies with facilities. The criteria can be modified to reflect changes in the administration's priorities.

STRENGTHS, WEAKNESSES, AND BEST OPPORTUNITIES FOR USE

The deferred maintenance scoring process combines an infrastructure for building condition data and inputs from agencies that manage the facilities. Experience in developing and using the process underscores the importance and challenge of defining the indicators and weights to reflect specific priorities.

EXHIBIT 1: PRIORITIZATION CRITERIA FOR DEFFERED MAINTENANCE

CRITERIA	RATING SCALE	DEFINING/CLARIFYING QUESTIONS
Tier 1 (weighting factor of 2.0)		
Degree of health/life Safety Impact	0-3 (none – highest)	How dangerous is the deficiency to occupants and visitors?
Urgency of Work	0-3 (none – highest)	How long has this deficiency been unaddressed? How quickly will things get worse left unaddressed?
Penalty Threat	0/1 (none/yes)	Does the deficiency put the facility out of regulatory compliance, or is there pending litigation regarding this deficiency ?
TIER 2 (WEIGHTING FACTOR OF 1.5)		
Component Priority	0-3 (none – highest)	How important is the deficient component to the overall site condition?
Shut Down Threat	0-3 (none – highest)	How important is correcting this deficiency to occupants' ability to carry out their core mission? Does the deficiency prevent use of the facility?
TIER 3 (WEIGHTING FACTOR OF 1.0)		
Future Plans for Site	0/1 (no/yes)	Is this facility/deficient component included in the long-range use plan for the site?
Ready to Proceed	0/1 (no/yes)	Is the project ready to proceed when funding is allocated?
Oversight Complete	-1/0 (no/NA or yes)	Have all of the required oversight documents been submitted?
Expended Previous Funds FY08	-1/0/1 (no/NA, yes)	Has the agency properly expended all deferred maintenance funds from previous fiscal years?
Expended Previous Funds FY09	-1/0/1 (no/NA, yes)	Has the agency properly expended all deferred maintenance funds from previous fiscal years?
Participating in Demand Response	-1/0/1 (no/NA/yes)	Is this agency eligible and enrolled in the Demand Response program?
TIER 4 (INFORMATIONAL PURPOSES)		
Additional information including costs estimate		

FACT SHEET | OECD Environmental Indicators

SPONSORING ORGANIZATION: Organisation for Economic Co-Operation and Development

REFERENCES:

Organisation for Economic Co-Operation and Development (OECD). "OECD Environmental Indicators: Development, Measurement, and Use." Reference Paper. Available at http://www.oecd.org/dataoecd/7/47/24993546.pdf.

Organisation for Economic Co-Operation and Development (OECD). "OECD Key Environmental Indicators, 2004". OECD Environment Directorate, 2004.

PROGRAM DESCRIPTION:

MOTIVATIONS

The Organisation for Economic Co-Operation and Development (OECD) has developed a set of environmental indicators in response to growing demands for reliable, transparent, and informative environmental data. These indicators are meant to improve countries' capacity to monitor and assess environmental conditions and trends, increase countries' accountability, and to evaluate if they are meeting domestic and international objectives. The indicators are also designed to promote cooperation and the exchange of information among countries, including non-member nations.

NUMBER OF INDICATORS, WHAT THEY COVER, AND WHY THEY WERE SELECTED

OECD has developed multiple sets of indicators. The Core Environmental Indicators are commonly agreed to by OECD countries and represent the main environmental concerns in those countries. There are approximately 50 indicators in this set.

The Key Environmental Indicators are a subset of the Core Environmental Indicators. These ten indicators are useful in informing the general public and providing policy makers with signals on major environmental issues.

The Sectoral Environmental Indicators, Indicators Derived from Environmental Accounting, and Decoupling Environmental Indicators are supplemental sets of indicators that provide more tailored environmental performance information.

OECD selected indicators using three basic criteria:

- **Policy Relevance and Utility for Users** – An indicator should be: representative of the country's environmental conditions, environmental pressures, and societal responses; simple and transparent; responsive to changes; and comparable to other nations.
- **Analytical Soundness –** An environmental indicator should be: theoretically well founded based on international standards or consensus and usable in economic models and forecasts.
- **Measurability –** Indicators should be selected where the data required is readily available, adequately documented, and updated regularly.

MEASUREMENT APPROACH:

DATA REQUIREMENTS TO COMPILE INDEX

Most core indicators are calculated using data collected by OECD from questionnaires on the state of the environment sent to national authorities. Data is checked for quality, treated, and harmonized to be compatible with the rest of the data set.

CONSTRUCTION

OECD does not compile its indicators into a specific index. It classifies the core indicators according to the Pressure-State-Response (PSR) model as indicators of environmental pressures, environmental conditions, or societal responses. OECD then groups the indicators into a set of environmental issues which reflect the major environmental challenges in OECD countries. OECD uses the PSR framework for other indicators, as well, but uses slightly different classifications.

USE AND RESULTS:

HISTORY OF USE

OECD has adopted the use of these indicators in its Country Environmental Performance Reviews. The international indicators from the OECD data sets are used in combination with specific national indicators and data in order to track trends over years and monitor changes in response to policy measures.

STRENGTHS, WEAKNESSES, AND BEST OPPORTUNITIES FOR USE

OECD's indicators are a useful tool for evaluating the environmental performance of individual countries. The indicators can be aggregated by sector or territory for more refined comparison.

The indicators are useful tool for evaluating environmental performance. While they are not designed to provide a full picture of environmental issues, they do help reveal trends and draw attention to phenomena or changes that require further analyses and possible action.

When using the OECD indicators, it is important to take into account the appropriate context for each county. While in some cases, it may make sense to compare two (or more) countries' index scores in absolute terms, it may also be helpful to compare these scores in combination with other factors that influence performance, such as GDP, population, or land area.

FACT SHEET | ONTARIO COMPLIANCE INDEX

SPONSORING ORGANIZATION: Ontario Ministry of the Environment

CONTACT: Rajeev Narang, Environmental Program Analyst, Sector Compliance Branch

REFERENCES: Ontario Ministry of the Environment. *Compliance Index(CI): Draft Report.* June 3, 2008.

Narang, Rajeev, Ontario Ministry of the Environment. "Compliance Index – Environmental Compliance/Performance Measure." Presented at the National Environmental Partnership Summit, Baltimore, MD. May 20, 2008.

PROGRAM DESCRIPTION: Ontario Ministry of the Environment's (MOE's) Compliance Index allows the agency to assess relative compliance within programs, industries, and sectors. It can be described as an outcome-based performance measure used to establish baseline compliance levels and measure changes through subsequent inspections.

MOTIVATIONS

MOE had two objectives in developing the index:

- To support compliance and performance measurement
- To target highly noncompliant facilities and sectors

STAKEHOLDER INVOLVEMENT

MOE had to make a number of key decisions while developing the index. First, MOE decided to target all environmental legislative violations (across all media and all enforceable sections. From there, the agency assigned weights to violations of each of approximately 1,300 legislative provisions. This was a time-intensive process that required substantial discussion between inspectors and other agency staff.

NUMBER OF INDICATORS, WHAT THEY COVER, AND WHY THEY WERE SELECTED

MOE's Compliance Index has two components: the legislative score and certificate of approval (CofA)[28] score. For legislative violations, there are approximately 1,300 separate provisions to score. The specific legislative and CofA provisions were chosen

[28] Any facility that releases emissions to the atmosphere, discharges contaminants to ground or surface water, provides potable water supplies, or stores, transports or disposes of waste must have a Certificate of Approval before it can operate lawfully. Each completed Certificate of Approval addresses matters that fall within the mandate of the ministry, focuses on site specific characteristics relevant to each proposal and contains enforceable requirements for each facility to ensure protection of human health and the natural environment. *Adapted from the Ontario Ministry of the Environment website: http://www.ene.gov.ca.en/business/cofa/index.php*

because they were enforceable provisions that would allow MOE to measure a facility's or sector's compliance with applicable requirements.

MEASUREMENT APPROACH:

DATA REQUIREMENTS TO COMPILE INDEX

The data for the index is gathered through facility inspections. Inspectors must first identify all potentially applicable legislative provisions and then gather data on whether the facility is in compliance with those provisions.

This information needed was already being collected in inspections via a checklist. MOE was able to use the same database to track compliance. The only change in the data collection process was in how the information was compiled and presented.

EQUATIONS USED

Both the legislative and CofA scores are the weighted sum of a facility's violations. The overall Compliance Index score is a sum of the legislative and CofA scores:

$$CI = \sum V_w + \sum C_w$$

where V_w = weighted legislative violation, and

C_w = weighted CofA conditions violation

WEIGHTING

Each provision violation is classified as one of four "contravention categories" (reporting and recordkeeping, operating standards, monitoring and sampling, or discharge/exceedence of limits) and assigned a corresponding weight of 1, 2, or 4. These weights represent the level of potential impact on the environment and human health, with higher weights representing a larger impact. (Note that there are only three weights but four categories.) The same process is used to score the CofA violations. MOE designed its index in this manner because it wanted both to identify the legislative violations and capture some level of consequence.

KEY ISSUES:

KEY STATISTICAL ISSUES, TOOLS AND TECHNIQUES

One issue with this pilot approach is that the weighting scheme applied to the different types of violations is relatively simplistic: there are only three possible weights. This scheme does not fully account for the variability of potential environmental and human health impacts of these violations. MOE is interested in enhancing the weighting process so that it can better capture differences in the potential impacts of the various types of violations.

In addition, MOE wants to address sectors with high levels of environmental risk, but at the same time, it wants to be able to use random samples of facilities to make inferences about the population of facilities in a sector. Essentially, MOE is using a targeted approach for determining which sectors to address and a statistically based random sample approach for which facilities to inspect in a sector.

USE AND RESULTS:

HISTORY OF USE

The Compliance Index is currently used within some divisions of MOE. MOE is working to expand its use to all the divisions within the agency. The index and its results have been well received but MOE sees the current index as a first pass and plans to make revisions and refinement.

STRENGTHS, WEAKNESSES; AND BEST OPPORTUNITIES FOR USE

MOE considers the greatest strength of the index that it provides an objective tool to measure how facilities are performing. Both the agency and regulated industries have recognized a need for more clarity and objectivity in compliance measurement.

However, MOE recognizes that the current index has some weaknesses. It is simplistic. Also, there has also been some confusion because higher scores do not correspond to higher risks – while the weights represent to some degree the significance of a violation, they do not consider the environmental risk that a specific case of non-compliance poses. MOE would like to clarify the level of significance of violations, add more categories (weights), and expand the indicators to capture the use of regulatory tools like permits and approvals.

FACT SHEET | Operational Risk Appraisal Scheme (Opra) - United Kingdom

SPONSORING ORGANIZATION: United Kingdom Environment Agency

CONTACT: David Pugh, Better Regulation Policy Manager, United Kingdom Environment Agency

WEBSITE AND REFERENCES:

http://environment-agency.gov.uk/opra

U.S. EPA, National Center for Environmental Innovation. "An In-Depth Look at the United Kingdom Integrated Permitting System." July 2008.

UK Environment Agency. "Spotlight on business – 10 years of improving the environment." July 2008.

PROGRAM DESCRIPTION:

The United Kingdom's Environment Agency (EA) has developed the Operational Risk Appraisal (Opra) as a tool for assessing the potential risk associated with facilities that can be used to help target agency resources to more effectively meet its objectives of environmental protection. Facilities applying for a permit must submit an Opra profile, which the EA then scores and uses in assessing the level of effort and resources required for permitting and monitoring compliance of the facility.

MOTIVATIONS

Opra was initiated in the 1990s through a doctoral thesis of an EA employee. The EA wanted a way to better define the risk of the facilities they regulate. Opra is based on the concept that once an agency can clearly define risk, it can more effectively target its efforts and resources to meet its objectives.

STAKEHOLDER INVOVLEMENT

When the EA developed the current version of Opra, it held a series of stakeholder workshops with businesses, conducted a pilot program to evaluate the scoring system within Opra, and engaged in a public consultation process to solicit feedback.

NUMBER OF INDICATORS, WHAT THEY COVER, AND WHY THEY WERE SELECTED

A facility's Opra score is based on five attributes: Complexity; Emissions; Location; Operator Performance; and Compliance Rating. The attributes consist of sets of indicators, or responses to objective questions, that signify the level of risk. Examples of indicators for each attribute include:

- **Complexity**
  - Type of activities occurring at the facility
  - Inventory of potentially hazardous materials used

- Size relative to sector

- **Emissions**
  - Type and quantity of potentially released substances
  - Media into which releases would take place
- **Location**
  - Proximity to human activity or residence
  - Status and sensitivity of the environment in the vicinity of the facility
- **Operator Performance**
  - Presence/absence of environmental management systems
  - Staff competence and training
  - Emergency planning
  - Enforcement history
- **Compliance Rating**
  - Instances of non-compliance with permit or license
  - Potential impact resulting from non-compliance

MEASUREMENT APPROACH:

DATA REQUIREMENTS TO COMPILE INDEX

When a facility submits its application for a permit from EA, it must also submit an Opra profile. The facility completes the information to score the first four attributes. EA has developed a spreadsheet tool that compiles the data and calculates the scores for each attribute. This tool is available on EA's website for facilities to complete. The agency completes the fifth attribute (compliance rating) based on its compliance classification scheme. Opra scores are updated by facilities when their operations change significantly enough to affect their ratings. The EA updates the compliance attribute on an annual basis.

WEIGHTING AND CONSTRUCTION

Each attribute of Opra is calculated differently. Answers to questions translate into scores that are in some cases weighted to reflect the environmental risk they imply. Although assigning scores and weights in some cases requires agency staff to make judgments, EA developed the scoring and weights to be as objective as possible.

The applicant's answers to the indicator questions are converted to an overall point score. This score is then used to create the facility's banded profile – a rating of A through E for each attribute, with A representing the lowest risk and E the highest risk. The banded profiles are used by the agency to determine compliance assessment priorities as well as the application, subsistence, and other fees the applicant will be required to pay.

KEY ISSUES:

KEY STATISTICAL ISSUES, TOOLS AND TECHNIQUES

In translating the numerical points score into a banded profile, EA had to decide upon cut points for the range of potential facility scores in each attribute. EA has based this scheme on the potential impacts for each of the attributes on policy decisions as to where the "norm" for facilities should fall within the banded scale for the attributes.

In some cases, facility scores for an attribute do not tend to fall in a normal distribution. For example, there are many small, less complex sites and fewer large, more complex sites. In addition, EA is starting to find that more and more facilities are falling into bands "A" and "B." While this reflects improvements in facility performance and operations, it does pose a potential future challenge in thinking about how to ensure the Opra system continues to be an effective tool in increasing environmental protection. For example, EA believes it might be appropriate to recalibrate the banded scores over time to better capture continuous facility improvement.

USE AND RESULTS:

HISTORY AND CURRENT USE

The first version of Opra considered only two factors: a pollution appraisal and operator performance. EA used descriptions provided by permit applicants to decide where they fit on a scale of these two attributes. EA updated Opra in response to concern that the initial system was too subjective. In the updated version, the agency has attempted to utilize more objective and specific indicators to support the final Opra score and resulting fees.

As noted above, Opra allows EA to evaluate which facilities pose the highest risks to the environment and human health, allowing the agency to target its resources and set objectives. Additionally, the banded profile scores are used to determine the cost of application, subsistence, and other fees for the facility. This system allows EA to recoup higher revenues from facilities that are likely to require the most resources from the agency.

EA also uses Opra scores to track trends in performance. The agency can track the change in banded scores and see whether its actions are helping to reduce the numbers of facilities with scores in the higher risk bands.

At the sector level, EA uses Opra to help to assess overall business performance. The distribution of facilities' operator performance banded scores is tracked over time to assess trends in performance. The scores identify sectors where additional EA attention may be warranted.

STRENGTHS, WEAKNESSES; AND BEST OPPORTUNITIES FOR USE

In addition to its utility in allocating resources and assessing fees, Opra scores are useful in other ways, such as:

- Providing a benchmark for the facility and the public to assess performance
- Targeting poorly performing facilities for compliance assessment

- Targeting additional areas for sector-wide environmental improvement
- Encouraging better facility performance by tying fees to environmental risk

Looking forward, EA continues to explore ways to make Opra more sensitive in differentiating between good and bad facility performance, especially in the operator performance attribute.

FACT SHEET | Operator's Safety Index (OSI)

SPONSORING ORGANIZATION: U.S. Department of the Interior, Minerals Management Service

CONTACT: Doug Slitor, Chief, Safety and Enforcement Brach, Minerals Management Service, Department of the Interior

REFERENCES: Slitor, Doug. "Measuring Safety and Compliance in the U.S. Offshore Oil and Gas Industry." *Society of Petroleum Engineers.* Paper prepared for presentation at the SPE International Conference on Health, Safety, and the Environment in Oil and Gas Exploration and Production held in Stavanger, Norway, June 26-28, 2000, SPE 61155.

PROGRAM DESCRIPTION: The U.S. Department of the Interior's (DOI) Mineral Management Service (MMS) has created an inspection program that periodically inspects offshore oil and gas operators for compliance with federal safety regulations. It compiles an operator's incidents of noncompliance (INCs) into an Operator's Safety Index (OSI) that accounts for operator compliance and accidents.

MOTIVATIONS

The DOI has been regulating the offshore oil and gas industry for over 50 years. Throughout this time period, DOI has promulgated several regulations designed to increase the safety of these offshore operations. The number of operators more than doubled from the mid-1970's to 1999, the number of facilities rose from 350 to nearly 4,000, and the number of wells drilled skyrocketed from 2,000 at the end of the 1950's to 38,000 by January of 2000. In addition, MMS was seeing a sharp rise in accidents and injuries at facilities during this time period. This rise in regulated operations and accidents, as well as the large number of applicable regulations (over 600), created a need for MMS to be able to normalize and compare operator compliance and accidents. MMS has also found a high correlation between instances of noncompliance and the probability of an accident at a facility. MMS wanted a way to compare the performance and safety of facilities with a wide variety of operations and size, so it created a unified scale to compare compliance and safety across operators over time. MMS also created the OSI to highlight operators with compliance or safety problems.

NUMBER OF INDICATORS, WHAT THEY COVER, AND WHY THEY WERE SELECTED

MMS transformed all of its requirements into a national list of over 600 "potential incidents of noncompliance" (PINCs), associated with the following 10 regulatory categories: abandonment, completions, drilling, environmental, general, hydrogen sulfide, pipelines, measurement, production, and workovers. Inspectors use a checklist of 600-plus requirements and issue "instances of noncompliance" (INCs) for any violation.

MEASUREMENT APPROACH:

DATA REQUIREMENTS TO COMPILE INDEX

All data used to compile the OSI is gathered through on-site facility inspections. MMS is required by current regulations to conduct at least one announced inspection per facility per year; MMS is also allowed to conduct periodic unannounced inspections throughout the year. In addition to information gathered about regulatory compliance, the OSI captures information about the severity of any accidents at the facility during the year. For each accident, MMS collects data about the injuries/fatalities, amount of spillage, type of event, and amount of damage.

EQUATIONS AND WEIGHTING USED

The OSI has two components: (1) instances of noncompliance and (2) accident severity.

(1) Instances of noncompliance are weighted by the type of the enforcement actions:

$$\frac{W(1)+C(2)+S(4)+CP(8)}{CI}$$

Where,

W = warning INCs;

C = component shut-in INCs;

S = facility shut-in INCs;

CP = INCs referred for a civil penalty; and

CI = total number of components inspected

The weights of the INCs (1, 2, 4, and 8) are a simple geometric progression and reflect MMS's increasing concern with the different types of noncompliance. They are not based on a statistical analysis.

(2) Accident severity is determined by combining a "severity value" for each component of an accident (i.e., injury/fatality level, amount of spillage, type of event, and property damage amount). Each component of an accident is assigned a value, ranging from 0 to 640, based on the extent of the damage or injuries resulting from it. These values were determined by industry expert opinion, values of previous accidents, and OSHA definitions (for injuries). For example, if one accident had minor injuries, with 50 barrels spilled, and $100,000 of property damages, the three corresponding severity values (0, 40, and 80) would be added together to produce the accident severity score (for that incident): 120.

$$\frac{AS}{CA}$$

Where,

AS = total accident severity; and

CA = total number of components available[29]

[29] The total number of components available (CA) represents the total number of components at the facility; whereas the total number of components inspected (CI) represents the total number of components included in the inspection. The CI could be less than the CA.

(3) The noncompliance and accident severity components are simply added together to produce the facility's OSI score:

$$OSI = \frac{W(1) + C(2) + S(4) + CP(8)}{CI} + \frac{AS}{CA}$$

KEY ISSUES:

KEY STATISTICAL ISSUES, TOOLS, AND TECHNIQUES

MMS considers this index to be an "indicator of operator compliance with recommended industry practices and Federal regulations;" it does not consider it a comprehensive measurement tool. In any facility evaluation or industry-wide comparison, it uses anecdotal information as well. The index provides a way to standardize and compare facilities that can vary greatly in size and operations.

At the time of the index's development, MMS held an open forum with the industry. There was concern expressed that the index was essentially double counting instances of noncompliance. If, for example, a facility received a warning INC and did not rectify the situation, that warning could turn into a civil penalty. If both the warning and the penalty occurred in the same year, the violation would be counted twice in the facility's OSI: once for the warning INC, and again for the civil action. However, MMS decided that if a facility received a warning that it did not rectify, the violation should be counted again. If a facility *did* resolve a violation for which it received a warning, that INC would only be counted once.

USE AND RESULTS:

PAST AND CURRENT USE

MMS began compiling this index in 1995 and uses the results for six primary purposes:

1. *MMS Safety Award for Excellence (SAFE)* – The OSI serves as the basis for nominating the finalists for MMS' SAFE awards, which are given out annually.
2. *Annual Performance Reviews* – Every year, MMS ranks all of the operators by their OSI score, and uses that ranking to help prioritize annual performance review meetings.
3. *Strategic Operator Performance Improvement* – At each facility's annual performance review, the OSI, along with other relevant information, is reviewed to determine a strategy to improve compliance and safety at the facility.
4. *Operator Disqualification* – Using the OSI, MMS separates operators into two categories: acceptable and unacceptable. Acceptable operators generally have an OSI score of less than 1.0.[30] Operator performance for unacceptable operators, including OSI scores, is examined for the last three years to determine an appropriate disciplinary action, including operator disqualification.
5. *Risk-Based Inspection Strategies* – MMS has found that compliance is one of the top predictors of future accidents at a facility; therefore, it has developed a risk-based strategy for prioritizing the scheduling of inspections. The OSI is one factor used in this process.

[30] 95 to 97 percent of facilities achieve an "acceptable" OSI score of less than 1.0.

6. *Government Performance and Results Act (GPRA) Reporting* – MMS uses the accident severity portion of the index as a performance measure to be included to the agency's GPRA reporting.

MMS is also seeing operators respond positively to the OSI. Operators are considering their OSI score as an important part of their internal monitoring process, and strive to be nominated for SAFE awards every year.

MMS also uses the index as a way to track performance throughout the industry, and to highlight trends (negative and positive) across facilities. Say, for example that the industry-wide OSI is unusually high in one year; MMS will go back and examine the facility-level scores to determine if the rise reflects a general problem in the industry or whether there are a few facilities with major problems.

FUTURE USE

MMS has issued a proposed rule requiring operators to develop and implement a Safety and Environmental Management System to address oil and gas operations in the Outer Continental Shelf. The system would consist of four elements: hazard analysis, management of change, operating procedures, and mechanical integrity. MMS found that the majority of accidents in the industry were related to operational and maintenance procedures or human error. As part of the new requirements, MMS would require operators to annually submit the number of hours worked for all company and contract employees during production, drilling, pipeline, and construction activities.[31] The number of hours worked would then be used as the "normalizing" factor for the accident severity component of the OSI. Essentially, the number of hours worked would replace the number of components at a facility because the agency believes that the number of hours worked better capture the facility's "exposure to events." MMS recognizes that once it switches over to this new measure, it will need to consider how to compare scores based on the revised index to the OSI from previous years; it has not yet resolved this issue.

[31] Department of the Interior, Mineral Management Service. "Safety and Environmental Management Systems for Outer Continental Shelf Oil and Gas Operations." Proposed Rule: 30 CFR Part 250, RIN 1010-AD15. Docket ID: MMS-2008-OMM-0003. FR Vol. 74, No. 115, Wednesday, June 17, 2009.