Research article

Behavioral patterns of environmental performance evaluation programs

Wanxin Li,*, Volker Mauerhofer

A Department of Public Policy, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong Special Administrative Region
b Department of Botany and Biodiversity Research, University of Vienna, Giessaufg. 28/5/34, 1050 Vienna, Austria

Abstract

During the past decades numerous environmental performance evaluation programs have been developed and implemented on different geographic scales. This paper develops a taxonomy of environmental management behavioral patterns in order to provide a practical comparison tool for environmental performance evaluation programs. Ten such programs purposively selected are mapped against the identified four behavioral patterns in the form of diagnosis, negotiation, learning, and socialization and learning. Overall, we found that schemes which serve to diagnose environmental abnormalities are mainly externally imposed and have been developed as a result of technical debates concerning data sources, methodology and ranking criteria. Learning oriented scheme is featured by processes through which free exchange of ideas, mutual and adaptive learning can occur. Scheme developed by higher authority for influencing behaviors of lower levels of government has been adopted by the evaluated to signal their excellent environmental performance. The socializing and learning classified evaluation schemes have incorporated dialogue, participation, and capacity building in program design. In conclusion we consider the ‘fitness for purpose’ of the various schemes, the merits of our analytical model and the future possibilities of fostering capacity building in the realm of wicked environmental challenges.

1. Introduction

Since the 1970s, environmental protection has become one of the core tasks of the government. Global, regional, and national policy makers have declared their ambition to improve the environment and pursue sustainable development driven by universally agreed principles and goals at the UN Conferences related to the environment, the Millennium Development Goals (especially goal No. 7 concerning environmental sustainability) and based thereon various implementations of the Agenda 21 cascaded down to national and local level. Measuring public service performance has stimulated an interest in using quantitative evaluation schemes in the public sector since the 1980s to improve efficiency (Downs and Larkey, 1986) and to enhance control and accountability of management behaviors, including environmental management (Halachmi, 2002; Simões and Marques, 2012).

In accordance with these broad developments a global trend of utilizing quantitative information to indicate the state of the environment and to evaluate efforts taken to improve it has emerged. This has occurred at different geo-political levels – global, regional and national – and according to different frameworks of reference – most notably, sustainable development (Mauerhofer, 2008), environmental performance management (Ewing et al., 2010) and quality of life (Ferreira and Moro, 2010).
Those efforts were intended to facilitate diagnosis of environmental problems and control agents, especially governments, who are capable of making a difference.

At the same time, the ‘wicked’ nature of many environmental problems questions the idea of control but calls for public managers to perform a role as collaborative capacity builder to facilitate learning, socialize environmental stakeholders and legitimize environmental actions (Hofstede, 1978, 1981). Weber and Khademian’s (2008) recent review identified three striking features that resonate strongly with the context of wicked environmental issues of concern here. First, wicked problems are unstructured in that the causes and effects are extremely difficult to identify and model. Second, wicked problems comprise multiple, overlapping and interconnected subsets of issues that cut across multiple policy domains and levels of government (Weber and Khademian, 2008). Finally, Weber and Khademian note that wicked problems are relentless.

In confronting wicked problems, where stakeholders are highly diverse and the output and effects of an intervention are highly uncertain, there have been growing calls amongst scholars to adopt “socializing and learning” management tools to create shared values and facilitate collaborative actions (Hajkowicz, 2005; Larson and Lach, 2008; O’Lear and Bingham, 2003, 2009; Ouchi, 1979; Rhodes and Murray 2007; van Elsacker et al., 2008; Zagonari, 2008). In this way a broader range of concerns and potential impacts can be considered beyond those of the immediate issue under scrutiny. However, the existing literature has mainly examined the scientific merit of some of the quantitative evaluation schemes (Benedetti et al., 2008; Finn et al., 2009; Gouveia et al., 2004; O’Lear, 1997, 1999; Roberts, 2006; Rogers and Louis, 2009) but overlooked the potential of using evaluation for facilitating learning among evaluators, environmental managers, and the public (Pollitt, 2013).

Our paper aims to develop a taxonomy of behavioral patterns in environmental management and to map against it ten purposively selected environmental performance evaluation programs at a national, regional, and global level, regarding their stimulated behaviors. The categorization of environmental management behaviors is built upon existing literature which we newly combine in an innovative matrix. Regarding the mapping, we mainly adopt a desktop research method of data collection, complemented with primary data collected from the first author’s participation in OECD environmental performance review for China and the assessment of provincial environmental performance in China by the YCEL and its partners. We found that all schemes reveal environmental performance information and help diagnose environmental abnormalities, while some have incorporated creative design features to go beyond diagnosis and to stimulate learning, influencing and bargaining, and/or socializing and learning. Diagnosis focused evaluation schemes hold a sharp dichotomy between evaluators and the evaluated and have been developed as a result of technical debates concerning data sources, methodology and ranking criteria. Learning oriented scheme is featured by an epistemic community, including both evaluators and the evaluated engaging in free exchange of ideas, mutual and adaptive learning. Scheme developed by higher authority offers political incentive for influencing behaviors of lower levels of government and has been adopted by the evaluated to signal their excellent environmental performance. Schemes developed by hub organizations of networks comprised of countries of different socio-economic and environmental conditions promote consensus building and environmental improvement through socialization and learning mechanisms such as dialogue, participation, and capacity building.

The next section shows the creation of the analytical tool categorizing management behaviors for addressing environmental challenges in a matrix form. Afterwards for the mapping working step, we first provide an overview of the ten evaluation schemes and then analyze their working procedures as well as behavioral patterns, and allocate the schemes within each classification quadrant of the matrix. Lastly, the paper discusses these results of this allocation and ends with conclusions.

2. Categorization of environmental management behaviors

Addressing environmental challenges requires both scientific knowledge and collective actions. Usually, it involves multiple stakeholders in problem identification and solution finding for pollution control, eco-conservation, and green transformation. Experts offer professional knowledge. Industrialists possess information on their production processes and associated environmental impacts that may or may not have been made public. Local residents understand practical know-how and local conditions. The government enacts and enforces regulations and policies to prevent and control environmental harm and to enable environmentally friendly behaviors (Marques and Simoes, 2008). Even more parties are involved if the geographical scale of an environmental challenge is at a regional or global level, e.g. climate change. It is understandable that those different actors pursue their self-interest and adopt different formulas in assessing costs and benefits associated with environmental problems and solutions.

Efforts in addressing environmental challenges may take different forms. It can be scientific research conducted in labs, by a group of scientists who share common goals. At the same time, collective actions are necessary among actors with diverse backgrounds, interests, and objectives. Information asymmetry is rampant among participants in those processes in terms of possession of scientific knowledge, level of individual contribution, and calculation of costs and benefits. Environmental impact of individual efforts may be uncertain or not immediately visible. How to manage those efforts and processes has become a tremendous challenge.

Scholars have studied the challenge of evaluating each individual’s contribution and distributing rewards in an equitable manner to ensure continuous cooperation within an organization. By whether output from individual efforts are measurable and whether members share common values or objectives, an organization can be managed by adopting one or more of the following three different mechanisms: markets, bureaucracies, and clan (Ouchi, 1979). Markets are applicable when effects of individual efforts are measurable, no matter whether or not members share common interests as they can negotiate on a fair share for oneself based on common knowledge of individual contribution. When individual output not measurable but members share common values and/or objectives, bureaucrats are a desirable form of management for organizing collective actions. When neither is output measurable, nor do individuals share common interests, it is better to manage as a clan by emphasizing socialization and creation of shared values (Ouchi, 1979). Thus, performance evaluation then can be used for diagnosing, bargaining, learning, and legitimizing in the following four scenarios where both output is measurable and objectives are shared by members, output is measurable but objectives are not common, output is not measurable but objectives are common, and neither is output measurable nor are objectives common, respectively (van Elsacker et al., 2008).

The above line of reasoning can be extended to the realm of environmental management. Addressing environmental challenges usually involves multiple organizations and individuals. Furthermore, not being given tasks to perform by the management, societal actors (usually) together with the government define and initiate desirable courses of actions. When societal actors share
common objectives, and causal links are established and solutions are readily available, “diagnosing” would suffice for addressing environmental challenges. Even with the same homogeneous group but when results from an intervention are hard to predict, “learning” has to take place because a trial and error process is necessary for solution finding. If societal actors are diverse in their backgrounds, values and beliefs, and/or objectives, “influencing and bargaining” can help reach consensus in situations where results from interventions are predictable. “Socializing and learning” are needed for creating shared values among diverse stakeholders and building their capacity to face up to the high uncertainty in outputs and effects of an intervention in an environmental challenge (Fig. 1). The relationship among these four distinct management mechanisms is neither uni-directional, nor is the distinction always a simple task and overlaps already occur by definition.

Fig. 1 combines in this way three essential aspects of environmental management, namely the procedural type represented by the four mechanisms, the stakeholder side in quantitative terms and the potential results alongside a gradient of uncertainty.

3. Data and methods

For analyzing behavioral patterns by design of environmental performance evaluation programs, we selected ten quantitative evaluation schemes with global, regional, and/or national coverage. All those programs are developed by reputable organizations: international/regional/national governing body, development assistance agency, and well-known independent research institute. We mainly adopt a desktop research method of data collection, complemented with primary data collected from the first author’s participation in OECD environmental performance review for China and the assessment of provincial environmental performance in China by the YCELP and its partners.

Overall, themes of the ten environmental performance evaluation schemes include sustainable development, environmental performance, and quality of life. The knowledge output takes the following different forms: indicators, index, review/assessment, and audit (Table 1).

By reviewing the features and evolution of those schemes, we analyzed with regard to the four behavioral patterns described above and allocated each scheme to one of the four matrix quadrants.

4. Design of and roles played by the ten selected quantitative evaluation schemes

The allocation of the ten selected quantitative evaluation schemes to the four roles of performance evaluation is reported in Fig. 2.

4.1. Diagnosing

Quantitative information is deemed necessary for evidence based decision making. The ESI/EPI, QECCUE, and CSDR were developed to bridge the policy-science gap by constructing indicator systems for diagnosing and identifying areas that need policy attention. Furthermore, government accountability can be enhanced by auditing that is result driven and places an emphasis on best value for money. The UA and QLA diagnose localities in their building livable cities and enhancing quality of life for residents. And the SDA diagnoses departments/agencies in their pursuit of sustainable development.

The ESI/EPI was initiated by Professor Daniel Esty at the YCELP in 2000, in collaboration with the CIESIN at Columbia University and the World Economic Forum (YCELP et al., 2000, 2001, 2002). The program aims to evolve from providing data for better environmental policy-making (YCELP et al., 2005) (2005 ESI report, page 2), to measuring proximity to explicit national environmental targets (YCELP et al., 2006: 1, 8), and to diagnosing environmental problems that may harm human health and/or the ecological system (YCELP et al., 2015). The number of countries covered increased from 56 in 2000 to 178 in 2014. Country ranking results are reported at the World Economic Forum every year.

Both the QECCUE and CSDR are limited to China. The QECCUE has been carried out by the Department of Pollution Control of the MEP since 1989. Moving away from qualitative to quantitative methods, the QECCUE was aimed to guide by scientific knowledge environmental protection work at a city level (MEP, 2008). Besides reporting to MEP, starting in 2002, participating cities were required to also publish in local newspapers the key environmental indicators. The MEP announced, in its “Annual national report on comprehensive control and management of urban environment” the names of the cities that are among the top ten in environmental quality, environmental management, and environmental infrastructure construction, respectively, and the top three that had made the most significant progress from the previous year (Li and Higgins, 2013).

To monitor and measure the implementation of China’s sustainability strategy, the Chinese Academy of Sciences has, since 1999, assessed and ranked the performance of the 31 regions in China and released the China Sustainable Development Strategy Report (CSDR) every February. The reports were distributed to the representatives of National People’s Congress members (NPC) and the Chinese People’s Political Consultative Conference members (CPPCC) in every March when they gather in Beijing for the annual meetings. Thus, it has drawn attention from the media and policymakers (Chinese Academy of Sciences (2014)).

A good quality of life is crucial for attracting and retaining a skilled labor force, businesses, students, tourists and, most of all, residents in a city. The Eurostat and the Directorate-General for Regional Policy of the European Commission thus co-developed an Urban Audit program in 1999 to promote enhanced quality of life amongst urban areas in EU member states. Built on the pilot study conducted in 58 cities, a meeting of ministers of the EU member countries on 8 October 2001 concluded that “statistical information on intra-city disparities is an indispensable base for further political action” (European Commission, 2004: 12).

In the UK, the Audit Commission, ODPM and DEFRA worked
Table 1
Summary of selected ten quantitative environmental evaluation schemes.*

<table>
<thead>
<tr>
<th>Evaluation scheme</th>
<th>Developer</th>
<th>Year of first release</th>
<th>Geo-political level</th>
<th>Theme</th>
<th>Knowledge output</th>
</tr>
</thead>
<tbody>
<tr>
<td>UN CSD ISD</td>
<td>Multiple parties supervised by the UN CSD</td>
<td>1996</td>
<td>Global</td>
<td>Sustainable development</td>
<td>Indicator</td>
</tr>
<tr>
<td>ESI/EPI*</td>
<td>YCELP, Yale University; CIESIN, Columbia University</td>
<td>2000</td>
<td>Global</td>
<td>Environmental performance</td>
<td>Index</td>
</tr>
<tr>
<td>OECD EPR</td>
<td>OECD Environment Directorate</td>
<td>1992</td>
<td>Membership based, regional – OECD</td>
<td>Environmental performance</td>
<td>Review</td>
</tr>
<tr>
<td>UA</td>
<td>Eurostat; Directorate-General for Regional Policy, European Commission</td>
<td>1999</td>
<td>Membership based, regional – EU</td>
<td>Quality of life</td>
<td>Audit</td>
</tr>
<tr>
<td>QEC CUE</td>
<td>Department of Pollution Control, MEP</td>
<td>1989</td>
<td>National – China</td>
<td>Control of urban environment</td>
<td>Quantitative examination</td>
</tr>
<tr>
<td>SDA</td>
<td>Commission for Sustainable Development</td>
<td>1995</td>
<td>National – Canada</td>
<td>Sustainable development</td>
<td>Audit</td>
</tr>
<tr>
<td>MCE</td>
<td>Department of Pollution Control, MEP</td>
<td>1997</td>
<td>National – China</td>
<td>Environmental protection</td>
<td>Award</td>
</tr>
<tr>
<td>CSDR</td>
<td>The Sustainable Development Strategy Study Group, Chinese Academy of Sciences</td>
<td>1999</td>
<td>National – China</td>
<td>Sustainable development</td>
<td>Indicator</td>
</tr>
<tr>
<td>QLA</td>
<td>Audit Commission; ODPM; DEFRA</td>
<td>2002</td>
<td>National – UK</td>
<td>Quality of life</td>
<td>Audit</td>
</tr>
</tbody>
</table>

* Programs listed from global to regional and national, chronically at each same level.

Fig. 2. Allocation of the evaluation schemes to the new matrix.

Together to develop QLA. The pilot set of quality of life indicators was released in 2002 and a revised version was released in 2005. Data on those indicators are available at local authority/local strategic partnership area level. This makes it possible for the Audit Commission to bring together robust, accurate data for each area to enable local comparisons. Thus, communicating and publishing quality of life indicators could raise the public awareness of the “livability” of their home places compared with other places and inform their actions for improving the local environment and tackling local issues (Audit Commission of the United Kingdom, 2005).

Under the 1995 Amendments to the Auditor General Act, the Commissioner of the Environment and Sustainable Development was appointed by the Auditor General of Canada to assist in carrying out the SDA. The SDA evaluates performance of the departments/agencies’ implementation of their sustainable development strategies. SDA reports are submitted to the House of Commons for review. A returning visit will occur two years after the audit to evaluate whether the deficiencies have been addressed and any improvement has been made. Individual interviews and symposium have both been used by the Commissioner’s Office for soliciting opinions from external stakeholders and experts and advisory committees for deciding what issues to audit. This gives the audits credibility and breadth (Office of the Auditor General of Canada and Commissioner of the Environment and Sustainable Development, 1999).

All the six programs share the following common features: expert driven, cross region comparison, and revelation of weak performance. Those programs, by design build in a sharp dichotomy between evaluators and the evaluated. Expert designed indicators and selected topics for audit give credibility to those programs. Results from comparison are presented in a report form and are made publicly available for both raising public awareness and naming and shaming laggards. In so doing, those programs can potentially bring to the attention of both policy-makers and the public of the evaluated country/region issues needed close attention and further actions but do not directly engage the evaluated to explore possibilities of change.

4.2. Learning

The OECD EPR falls in the upper left quadrant and works through “learning” to facilitate informed decision-making. In 1991, the OECD Environment Ministers, OECD Council, and Heads of State and Governments at their meetings endorsed the Group on Environmental Performance (GEP) to develop the EPR. The principal aim of the OECD’s EPR is to help Member countries improve their individual and collective performances in environmental management. Modeled after the well-known OECD Economic Survey, the EPR is grounded in the basic OECD function to review trends, policies and countries’ performance (OECD, 1997: 6).

The review processes include preparation, review mission, peer review by the GEP, publication, and follow-up and monitoring. It begins with the Secretariat consulting with the reviewed country to formulate an outline of the review. Then the Secretariat assembles a review team which typically includes experts from three reviewing countries, Environment Directorate staff and prominent consultants, and often an observer (e.g. from a different OECD directorate, from the UN-ECE). A set of discussion themes is prepared for each review for use as a kind of agenda during the team mission. It is circulated to participants in the country being reviewed, a month before the start of the visit. On the review mission, the expert team meets with government and non-government representatives of the country under review, including industry, trade unions, NGOs, experts and local government representatives. Four to six months after the review mission, a review report will be compiled and the
Secretariat will harmonize and consolidate it. The report is then sent to all capitals four to six weeks before the GEP peer review meeting. Examining a country review report takes up a full day. The reviewing countries take a lead in opening the debate with all countries participating in it. No minutes are taken at the meeting to allow free exchange of views. Besides the reviewed countries benefit from the discussions, reviewers can also bring the lessons back to their home countries. Publication of the completed report under the responsibility of the Secretary-General constitutes the last step of the review process. The reviewed countries are expected to give feedback on the use they have made of the OECD performance reviews in the form of either formal “government responses” or informal oral reports to the GEP (at its subsequent meetings), and will also come with the second cycle of OECD reviews.

The OECD EPR has, beyond an informational approach, built an epistemic community to facilitate mutual learning among OECD Member countries. The reviewer and reviewed countries engage in the open discussions throughout the peer review process. Particularly, the closed door and non-minute taking full-day examination of a country review report and follow up report by the reviewed country encourage a pragmatic approach to collectively search for solutions to environmental challenges as well as promote accountability.

4.3. Influencing and bargaining

The MCE scheme was first proposed in 1997 in the “Ninth Five-year Plan for Environmental Protection and the 2010 Long-Term Goals” by the then State Environmental Protection Administration (upgraded to MEP in 2008). It was built on the QECCUE and added indicators on economy and society to make sure the model cities are not only environmentally friendly but also economically prosperous and socially harmonious. Starting in 2007, the MEP requires applicant cities to also report the level of public satisfaction with city environmental management. The MCE is a voluntary scheme and only those cities that have passed the thresholds set by MEP on the indicators are eligible to apply.

Applicant cities first submit an application to the MEP together with their plans for constructing a MCE. After the eligibility check, the MEP will conduct on site examination and publish the results in specified local newspapers for at least 10 days for public comment. Then the title of a MCE will be granted at a formal meeting held by the MEP. All the MCEs have to pass the returning checks by the MEP to maintain their status.

The MCE in China provides political incentive to influence local governments on their choices of development strategies. By defining what makes a desirable urban jurisdiction and publicly acknowledging the ones that have met the criteria, the MEP pushed environmental agenda onto local governments that value the title of MCE. Local political ecology may change with the construction of a MCE because environmental agencies may be better empowered to advance environmental interests. Thus, MCE becomes an influencing and bargaining mechanism that impacts the interplay between different agencies with sometimes conflicting goals in local government (Li and Higgins, 2013).

4.4. Socializing and learning

The UN CSD ISD and GMS EPA fall clearly in the upper right quadrant and work through the “socializing and learning” mechanisms to build capacity of participating countries collaboratively. Since 1992, supervised by the UN CSD, government officials, experts, and representatives of NGOs have worked together to develop a set of indicators for sustainable development after the Earth Summit. It was recognized that environmental components of sustainable development are among the most difficult to adequately reflect in national level indicators and, consequently, the area of greatest need (UN Economic and Social Council, 1995). Thus, the ISD indicators developed were intended to provide a reference, or sample set, for use by countries around the world to track progress toward nationally-defined goals, in particular, and sustainable development, in general. Discussions and workshops held on ISD offer opportunities for socialization and learning. In the 1996 workshop alone, participants include 30 more UN agencies, national governments, 45 more experts from international organizations, academic institutions, NGOs and 100 more individual experts trying to build consensus on the thematic framework, guidelines, methodology sheets and testing, evaluating, and revising the indicators. Even more people engaged throughout the 21 workshops from 1995 to 2009. During the processes, ISD has been revised based on country experiences. Furthermore, the collaborating agencies have agreed to incorporate these indicators in relevant capacity-building activities and inter-agency cooperation to ensure coherence of ISD with other indicator sets such as the MDGs (United Nations Development Programme, 2012: 7);

Furthermore, a “twinning” strategy was adopted for pairing participating countries to facilitate sharing of information and experiences (e.g., South Africa and Finland) or for one country to provide technical and financial support to another country (France and Tunisia) (United Nations, 2007: 8).

The six CMS countries, Cambodia, Lao PDR, Myanmar, Viet Nam, and Yunnan province of China started the GMS Economic Cooperation Program in 1992, under which a Working Group on Environment was formed in 1995. Phase Two of the technical assistance project directed by the Working Group, Strategic Environment Framework funded by the ADB started to include EPA from 2003 to 2005. Consultation workshops are held in each country and at the sub-regional level for each country to identify six to seven priority environmental issues for initial EPA exercise, and to select at least one indicator for each issue (GMS Environment Operations Center, 2008b). Furthermore, in July 2005, GMS Core Environment Program was endorsed at the Second GMS Summit. Its implementation arm, GMS Environment Operations Center funded by the ADB was inaugurated in 2006, which is responsible for mainstreaming EPA through extensive training and capacity building and further promoting wider application of EPA at a sub-national level. It is hoped EPA will become a routine activity and part of the set of management tools that will be consistently applied in each GMS country (GMS Environment Operations Center, 2008a: 39). To assist implementation of the Core Environment Program, a high level technical advisory panel as well as a regional GMS university network of centers concerned with environmental management have been formed (GMS Environment Operations Center, 2008b: 7).

Thus, directed by hub organizations of a global and/or regional network of countries with diverse socio-economic and environmental conditions, both the UN CSD ISD and GMS EPA set the stage for multiple stakeholders to participate, socialize and learn from experts as well as from each other. Both programs have gone beyond diagnosis and have facilitated creation of shared values, consensus building, and capacity building for addressing environmental challenges by individual country/region as well as collectively. Especially with the UN CSD ISD, member countries were expected to develop their own indicator systems based on the ground work laid down by the UN, which exhibited the results of socializing and learning.
5. Discussion

Environmental challenges are of major societal concern in their own right and also feature as a subset of issues to be addressed in the pursuit of a better quality of life and sustainable development. They are wicked problems that call for innovative management tools for consensus building and solution finding. Multiple stakeholders, international and supranational organizations, national and local governments, and academics have been experimenting with quantitative environmental evaluation schemes for many years now. In addition to detailing the key features, content and evolution of the ten purposefully selected, this paper considered the extent to which each of the schemes, as historically configured, open up the possibility of ‘diagnosing’, ‘influencing and bargaining’, ‘learning’ and ‘socializing and learning’ and classified them accordingly.

A majority (six) of the ten evaluation schemes we considered fall into the ‘diagnosing’ category for revealing weak performance, regardless the evaluated being willing to participate or not (Kushman et al., 2013). In a number of instances they stimulated efforts amongst some of the low ranking participants to improve (Kushman et al., 2013). In a number of instances they stimulated interventions. Furthermore, using data collected by the UA, the Urban Atlas, a part of the Global Monitoring of Environment and Security (GMES) service funded by the European Regional Development Fund, was first developed in 2009 for 185 cities. Commission Vice-President Günter Verheugen, responsible for Enterprise and Industry, and Commissioner Danuta Hübner, responsible for Regional Policy, said: “Through the use of European space technology we will open up possibilities for mutual learning from a land use perspective and help cities to make more informed investment decisions” (European Commission, 2009).

Thus, diagnosis exercises offer potential for the evaluated entities to learn and improve, if intermediate organizations can effectively engage the evaluated. However, it remains a big challenge for program developers how to better realize this potential by constructing indicators that would have scientific rigor as well as incentivize learning and improvement.

In contrast, the OECD EPR does not name or shame weak performers by comparing and ranking member countries’ environmental performance. Instead, it builds an epistemic community and fully adopts a peer review process comprising mostly a revolving lesson-learning and guidance activity. Although the quantity of information arising from the OECD EPR process is likely to accumulate quite rapidly the nature of its dissemination is such that its magnitude will for the foreseeable future be confined to a relatively small and slowly evolving set of substantive considerations (Lehtonen, 2009).

The MCE in China offers political incentive to enhance influence potential. This organizing outcome reflects its pre-fabricated structure in terms of goal orientation and empowerment. By defining what makes a desirable urban jurisdiction the MCE scheme limits the potential for lesson learning. However, it instead provides a ‘carrot’ for interested parties to gain more influence in the sphere of activity facilitated by its endorsement.

Lastly, the hub organizations of the UN CSD ISD and GMS EPA set the stage for multiple stakeholders to participate, socialize and learn from experts as well as from each other. One of the key features that distinguishes this subset of evaluation schemes is the evaluator directly engage the evaluated countries with diverse socio-economic and environmental conditions in workshop and networking activities. In the case of the UN CSD a dialectic approach was adopted which involved rationalizing an initial set of indicators to a smaller more manageable number. In contrast, the style of socialization and learning in GMS EPA related more to the processes of issue identification and knowledge transfer.

6. Conclusion

Although this paper has been able to broadly classify each of the schemes, by its function according to a framework of four management mechanisms it has not been possible at this stage to support a definite conclusion on which scheme is desirable or undesirable for addressing environmental challenges. All schemes set off from diagnosis but some are able to incorporate creative measures to perform functions of learning, influencing and bargaining, and socializing and learning in addressing environmental challenges and bringing about change. Credible information and scientific knowledge alone cannot solve wicked environmental problems but can set the basis for negotiation, learning, and socialization among diverse stakeholders. Complemented by intermediate organizations, diagnosis focused schemes can facilitate the evaluated, especially the laggards to learn, to improve, and to reach the norm.

Thus, this study suggests a system approach to environmental managers and policymakers in adopting evaluation programs for addressing wicked environmental challenges. It is important to take into consideration the nature of the issue at focus described by diversity of stakeholders and uncertainty in outcomes of interventions. Furthermore, it is important to not only focus on the technical aspects of evaluation but also gauge necessary conditions to transform the evaluation into an agent of change by reaching out to intermediate organizations, and/or building epistemic community, and/or establishing channels of dialogue and collaboration. In this way, the scientifically rigorous evaluation results will enable formulation of policy instruments such as informational, economic, and/or political incentive mechanisms for advancing the environmental agenda and achieving environmental improvement on the ground.

This study has left the following questions for future research. Under what kind of circumstances, an evaluation program can maintain high scientific rigor and neutrality without holding a sharp dichotomy between evaluators and the evaluated? What kind of intermediate organizations are effective in transforming a diagnosis-focused evaluation program into an agent of change? How to best combine one or more of the four types of evaluation schemes with policy instruments that can fit the policymaking process and political culture in different contexts and geographical scales for delivering environmental improvement? For now, we hope at least that this paper has helped to open up the debate, and can better generally inform environmental managers and stakeholders in terms of mechanisms to address environmental challenges and how to categorize them by means of different behavioral patterns through an innovative matrix.

Acknowledgements

The authors thank Ian Scott, Paul Higgins and the three anonymous reviewers for very helpful comments. Financial support from the Research Grants Council of Hong Kong (project 9042115) is gratefully acknowledged.

References

GMS Environment Operations Center, 2008b. GMS Core Environment Program