# Japanese Journal of Evaluation Studies

## Vol. 9, No. 2, September 2009

### English Version

## CONTENTS

**Japanese Journal of Evaluation Studies is published twice a year in Japanese and once or twice a year in English.**

©Japan Evaluation Society

# Japanese Journal of Evaluation Studies

## Vol. 9, No. 2, September 2009
### English Version

## CONTENTS

【Article】

# An Empirical Study of the Methodology for Assessing Social Capacity : The Case of Urban Air Quality Management

Kazuma Murakami

Mitsubishi UFJ Research & Consulting Co. Ltd.

kazuma@murc.jp

Shunji Matsuoka

Graduate School of Asia-Pacific Studies, Waseda University

smatsu@waseda.jp

## Abstract

This study proposes a methodology for assessing social capacity for environmental management (SCEM) which contributes to urban air quality improvement by identifying the actors and factors constituting the social capacity explicitly. To do so, a confirmatory factor analysis and a structural equation model are applied to estimate the effects of social actors (i.e., government, firms, and citizens) and factors (i.e., policies & measures, resource management, and knowledge & technology) on urban air quality management in 2 major cities (i.e., Osaka, Kitakyushu) in Japan. Using a panel data set for years between 1971 and 2000, a new evaluation method for capacity development is proposed called a "Actor-Factor Matrix evaluation" by using the results obtained in the analyses. Our results also show the difference of structure in SCEM and contribution to urban air quality by each city.

## Keywords

Social Capacity for Environmental Management, Confirmatory Factor Analysis,
Structural Equation Modeling, Air Quality Management

## 1. Background and Purpose

'Environmental Governance' (Matsushita, 2002), that advances environmental management through the appropriate participation of governments, firms, citizens, and society as a whole. Given this, it is necessary to create as much capacity as possible for governments, firms and citizens to fill their roles as the bearers of responsibility for the environment as a common public good and provider of services.

Social Capacity for Environmental Management (SCEM) is the total capacity of society as a whole, links to the efficient and effective implementation of environmental policies and international environmental cooperation (Matsuoka et al., 2004). This overall capacity is comprised of the capacities of governments, firms and citizens to cope with environmental problems and the mutual relationships created between these actors. Therefore SCEM is a concept that considers the role-sharing of the actors involved in solving environmental problems, and can be used to develop methods for assessing the capacity required to fulfill those roles.

In regards to the environmental management capacity debated by the OECD/DAC (1999) and others, Janicke et al. (1997) and Weidner et al. (2002) have tried to establish that environmental capacity includes the capacities of actors and jurisdictions, their mutual relationships and the framework conditions (awareness and information, government and institutions, economy and technology) that influence them. The actors include governmental organizations and non-governmental organizations (environment protection groups, media and corporations that are proactive environmentally). However, these are taken only as filling roles in promoting changes in the behavior of pollution emitters. Victims and polluters are not specified. Similarly, as environmental problems have been becoming more complex recently, inquiries that are based on a limited selection of actors and set roles, such as only considering industrial pollution, are insufficient. Moreover, the settings for the capacity elements that are the subjects of evaluation as the capacity standards for every actor are defined narrowly. For example, the capacity element of governmental organizations is expressed simply as organizational structure development stages on a national level. In order, the levels are "no environmental ministry," "isolated environmental ministry with low position," "environmental departments established even by regional local governments," "environmental sections established in most government divisions," "environmental planning institutions established." Thus, issues related to actors and capacity elements remain unresolved.

In Murakami and Matsuoka (2008), exploratory factor analysis (EFA) as used to identify systematically the elements that form government air pollution countermeasure capacities. The contribution of these capacity elements to air quality improvement was empirically confirmed, resulting in standards for capacity assessment shown in Table 1. However, this analysis was limited to the capacities of governments and disregarded the capacity formation of firms and civil society[1], which are also relevant social actors. Still, the capacity element classifications shown in Table 1 for the policy-countermeasure and knowledge cycles are, to some extent, classifications based on ordinary standards, so they are also applicable in determining the capacity formation of firms and citizens.

In this article, we apply the government capacity assessment framework in Table 1 to empirically determine the capacity formation related to air pollution countermeasures of firms and citizens in addition to governments to clarify their roles in the formation of the SCEM that contributes to air quality improvement. Through this, we are able to develop a capacity assessment framework that can evaluate SCEM and each actor's environmental management capacity level from observed data.

In this article, section 2 details analysis subjects, analytical methods and data, while section 3 describes analysis results. Finally, in section 4, we show a conclusion.

## 2. Subjects and Methods of Analysis

### 2.1 Setting analysis subjects

We choose two air pollution substances with different emissions sources and degrees of solution difficulty as the subjects of our analysis. In addition to sulfur dioxide ($SO_2$), which is a typical industrial air pollution caused by factories, workplaces and other fixed emissions sources that were the main subjects of Murakami and Matsuoka (2008), we looked at nitrogen dioxide ($NO_2$), a typical substance emitted from not only fixed emission sources but also automobiles and other mobile emissions sources that contributes significantly to air pollution in the urban living environment[2].

As in Murakami and Matsuoka (2008), we chose the time from when environmental policies were actually implemented and results began to be apparent for the analysis period. In this case, we chose the 30-year period from 1971 to 2000[3]. According to Harashima and Morita (1995), this was the nascent period for environmental policies

## Table 1  Actor-Factor Matrix (Air quality management)

| | | Factor | | |
|---|---|---|---|---|
| | Process | Execution of policies and countermeasures (knowledge implementation ) | Evaluation of policies and countermeasures (knowledge reproduction) / Formation of policies and countermeasures (knowledge emergence and use) | Setting policy and countermeasure issues (knowledge creation and accumulation) |
| | Capacity | P: Policy and countermeasure execution capacities (policy & measure) | R: Environmental policy and countermeasure resource management capacities (resource management) | K:  Knowledge, information and technology provision capacities (knowledge & technology) |
| Actor | G | ○Regulatory methods<br>• Creation and application of air pollution countermeasure legal regulations<br>• Setting and enforcement of environmental and emission standards<br>• Creation and application of air pollution countermeasure ordinances and fundamental plans<br>• Observation through onsite inspection, etc.<br>○Market based methods<br>• Formation and application of environmental taxes, charges and subsidy payment systems<br>○Autonomous methods<br>• Formation of pollution prevention agreements (without legal basis)<br>• Promotion of environmental study and education | ○Funds, Budgets<br>• Expansion of air pollution countermeasure budgets<br>○Personnel and organization<br>• Establishment of environment‐related divisions<br>• Establishment of air pollution countermeasure organizations (committees, councils, business and citizen assemblies)<br>• Increase of air pollution countermeasure division staff<br>○Facilities and equipment, etc.<br>• Arrangement of air pollution monitoring systems<br>• Installation of air pollution warning equipment and arrangement of information systems | ○Investigation and research<br>• Research on air pollution causes, mechanisms etc.<br>• Development of air pollution countermeasure technology and know‐how accumulation<br>• Policy research on air pollution countermeasures<br>○Information disclosure and sharing<br>• Disclosure of air quality conditions and pollution countermeasure information<br>• Implementation of staff education and training |
| | F | ○Regulatory methods<br>• Observance of air pollution countermeasure laws and regulations<br>• Observance of environmental and emission standards<br>• Compliance with air pollution countermeasure ordinances and fundamental plans<br>○Market based methods<br>• Countermeasures through the use of subsidy payment systems, etc.<br>○Autonomous methods<br>• Formation of pollution prevention agreements (without legal basis)<br>• Environmental load reduction through the entire business process including financial service production<br>• Acquisition of ISO14001 certifications, introduction of ESCO projects, etc. | ○Funds, Budgets<br>• Expansion of air pollution countermeasure budgets<br>○Personnel and organization<br>• Establishment of environment‐related divisions<br>• Increase of environment‐related division staff<br>• Increase numbers of environmental managers and pollution prevention managers<br>○Facilities and equipment, etc.<br>• Arrangement of in‐house monitoring systems<br>• Arrangement of warning equipment and information systems | ○Investigation and research<br>• In‐house monitoring of factories and workplaces<br>• Development of air pollution countermeasure technology and accumulation of expertise<br>○Information disclosure and sharing<br>• Creation and disclosure of environmental reports and environmental accounting<br>• Implementation of staff education and training |
| | C | ○Regulatory methods<br>• Observance of air pollution countermeasure laws and regulations (field burning, etc.)<br>○Market based methods<br>• Countermeasures that make use of subsidy payment systems, etc. (NGO, NPO)<br>○Autonomous methods<br>• Complaints, requests and lobbying<br>• Transition to lifestyles that conserve energy and resources, including more ecological automobile use, public transportation use, etc.<br>• Green purchasing and environmental funds | ○Funds, Budgets<br>• Expansion of environmental countermeasure budgets<br>○Personnel and organization<br>• Increase of the number of environmental countermeasure personnel (NGO, NPO)<br>• Participation in NGO and NPO activities<br>• Participation in environmental events, etc.<br>○Facilities and equipment, etc.<br>• Acquisition of facilities and equipment related to environmental countermeasures (NGO, NPO)<br>• Introduction of environmentally friendly products (energy conservation, new energy devices) | ○Investigation and research<br>• Investigation and research (NGO, NPO)<br>• Observation and monitoring (NGO, NPO)<br>○Information disclosure and sharing<br>• Understanding air pollution status<br>• Implementation of environmental education and training |

Source: The Authors

(1965～), and, according to Matsuoka and Kuchiki (2003), this was the period when the social capacity for environmental management began to function significantly (1970～). Environmental policies started to be implemented seriously, and their results began to become apparent during these decades. Around the start of this period, according to data from long-term continuous monitoring stations around Japan, annual average $SO_2$ concentration peaked in 1967, and $NO_2$ concentration peaked in 1971 before starting to decline.

For our subject cities, as in Murakami and Matsuoka (2008), we choose Osaka City and Kitakyushu City because they have different urban structures and they are both ordinance-designated cities that have specified authority over environmental policies. These two cities have differences in fixed emission sources as shown in Figure 1[4]. In addition, for mobile emission sources, the numbers of automobiles, the average traffic volumes and the average congestion levels (ordinary road total) are shown in Table 2. As a result, differences in annual average $NO_2$ concentration values in long-term continuous monitoring of automobile emissions occurred as shown in Figure 2.

Figure 1  The Numbers of Manufacturing Establishments (left), Share of Value of Manufactured Goods
Shipments (manufacturing establishments with 300 or more employees)(right)
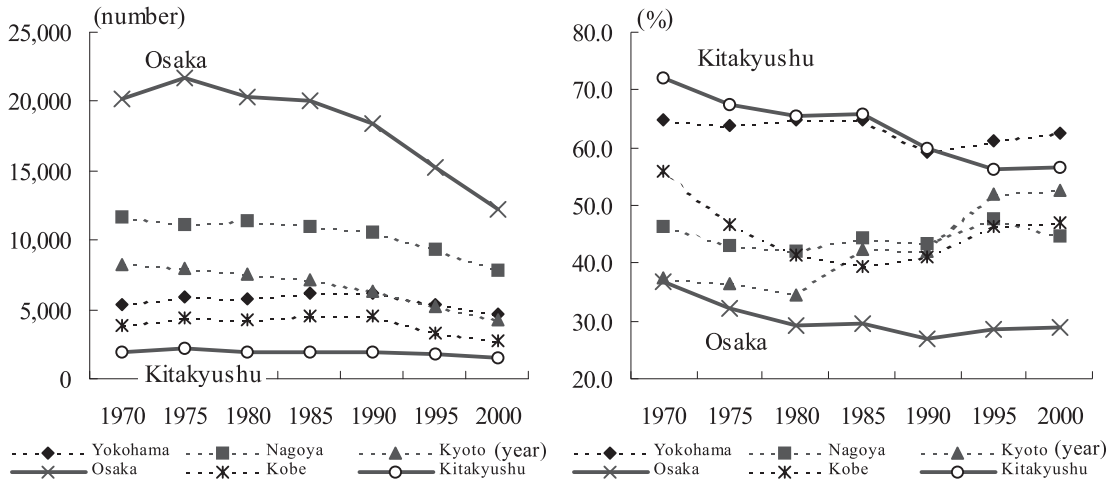
Furthermore, in Osaka Prefecture, a government designated area where the total NOx emissions volume is regulated that includes Osaka City, the automobile emissions ratio of the total NOx emissions amount was, according to estimates of the Japanese Ministry of the Environment, 47 percent in 1985 (Japan Society of Air Pollution, 1993) and 51 percent in 1997 (Osaka Prefectural Government, 2003). In contrast, in Kitakyushu City, the rate in 1983 was only 16 percent (Kitakyushu City, 1998). From this, we can see that these two cities with their different urban structures also have different $SO_2$ and $NO_2$ emissions structures and countermeasures (priorities) (Figure 3)[5].

With Osaka City and Kitakyushu City and their differing urban structures and air pollution countermeasures as our subjects, we investigated their structures of SCEM related to air pollution countermeasures and the influences on air quality improvement. Then, through comparison, we developed conclusions specific to each city and common to both.

Table 2  The Numbers of Automobiles, the Average Traffic Volumes and the Average Congestion Levels

|      | Numbers of automobiles | | DID・the average traffic volumes | | DID・the average congestion levels | |
| --- | --- | --- | --- | --- | --- | --- |
|      | Osaka | Kitakyushu | Osaka | Kitakyushu | Osaka | Kitakyushu |
| 1971 | 603,547 | 176,515 | 29,951 | 17,058 | 0.95 | 0.81 |
| 1980 | 718,755 | 307,867 | 24,981 | 17,222 | 1.04 | 1.00 |
| 1990 | 939,728 | 447,254 | 25,620 | 18,880 | 1.36 | 1.11 |
| 1999 | 913,285 | 552,573 | 23,799 | 19,578 | 1.07 | 1.09 |

Note: DID-Densely Inhabited District
The average traffic volumes =sum (traffic×road length) / total road length
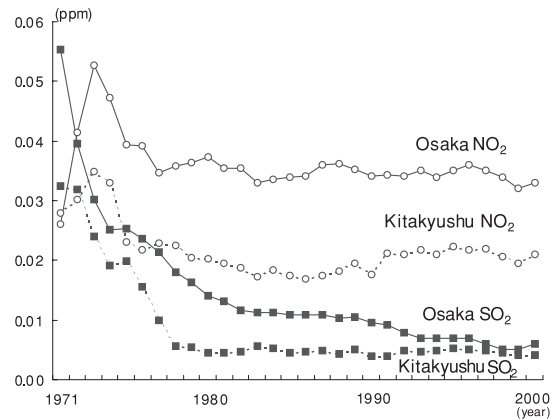The average congestion levels =traffic / traffic capacity
Source: Road Traffic Census / Ministry of Land, Infrastructure and Transport

### Figure 2  NO$_2$ Concentration (Automobile)
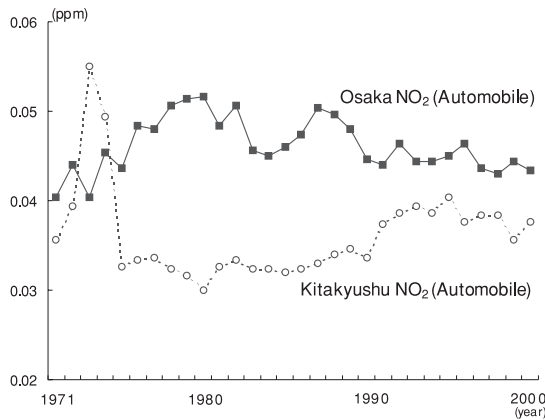


### Figure 3  SO$_2$ and NO$_2$ Concentration



Source: Air Pollution in Japan/Ministry of Environment

Source: Air Pollution in Japan/Ministry of Environment

## 2.2 Analytical methods

We sought to clarify the structures for social capacity for environmental management expressed by the environmental management capacities of governments, firms and citizens. Then, we made a formula for the causal relationship between SCEM and air quality improvement, and through verification, could show the structures of SCEM for air quality improvement. We conducted a two-stage empirical analysis to achieve this.

### 2.2.1 Analysis 1: verification of capacity formation

Through multi-group simultaneous analysis using a confirmatory factor analysis (CFA) model, we clarified the SCEM structures. Factor analysis[6] is a method for identifying latent variables that cannot be directly observed from the observed series of variables (Yanai *et al.*, 1990). In this case, we derived each actor's environmental management capacity and social capacity for environmental management, which are latent and cannot be directly observed, from each actor's observable activity level in relation to air pollution countermeasures. In addition, multi-group simultaneous analysis[7] is a method that investigates whether similar factor structures can be assumed across multiple populations. In this case, for both Osaka and Kitakyushu, to verify configural invariance, we assumed equivalent structures for the distribution of observed and latent variables (consistency of path position) and investigated both populations simultaneously to show the validity of the hypothesized capacity formation.

### 2.2.2 Analysis 2: verification of the causal relationship between capacity and air quality

We used a structural equation model (SEM)[8], which is a statistical approach that introduces latent variables that cannot be directly observed and identifies the causal relationship between these latent variables and observed variables in order to understand social and natural phenomenon (Kano and Miura, 2002). In this case, we formulated the social capacity for environmental management estimated in analysis 1 and the causal relationship with air quality improvement for each city. Doing this, we were able to verify the contributions of social capacity for environmental management to air quality improvement.

**2.3 Data**

As shown in Table 1, each actor's environmental management capacity can be expressed using 3 factors that correspond to policy-countermeasure cycle and knowledge cycle classifications. From these, we selected data according to the 3 factors classifications for the series of phenomenon related to the air pollution countermeasures of each actor's fundamental social role. We assumed these roles to be policies for governments, production for firms, and consumption for citizens. In short, the relationship between each actor's environmental management capacity and the "outcomes" of $SO_2$ and $NO_2$ concentration improvements can be expressed by the 3 data for "input" (K, R) and "output" (P) that indicate the levels of air pollution countermeasure achievements (Table 3).

Moreover, each actor's data were processed by different units that we then unified to create indexes. For governments, this was per city and environmental institute, for firms it was per workplace, and for citizens it was per household.

**2.3.1 Government environmental management capacity (G_cap)**

For governments, appropriate implementation of air pollution policies (G_P) can be set as the output of policies of air pollution countermeasures. Furthermore, behind these achievements is the scientific research (G_K) that is the basis for the policies and the organizational structures involved in making policies from that scientific research (G_R)[9]. Thus, a policy process series of G_K to G_R to G_P exists for air pollution countermeasures. In this case, the capacity elements that correspond to these can be represented by the 3 items under G_cap in Table 3.

Regarding environmental governance, Matsushita (2002) has expressed the need to divide the policy process into the setting of issues, policy development and policy implementation, and that there are gaps between the reality of environmental problems, the comprehension of them as governmental issues, policy setting and actual policy implementation. This indicates a need for investigation by policy process classification in order to grasp the causes of these gaps[10]. Thus, setting scientific research, the human resources related to policy creation, and policy implementation as individual capacity elements that correspond to policy process classifications is appropriate.

Table 3  Data

| | | Data | Source |
|---|---|---|---|
| G_cap (Government) | G_K | Number of academic articles about research results from city environmental science laboratories | Report of Kitakyushu City Institute of Environmental Sciences |
| | G_R | Number of city environmental division and environmental science laboratory staff members X average number of years employed | Annual Report of Osaka City Institute of Public Health and Environmental Sciences |
| | | | Environment in Kitakyushu City |
| | G_P | Number of air pollution inspections of factories, workplaces and other facilities by city divisions and environmental science laboratories | White Paper of Environment Pollution in Osaka City |
| | | | The Survey on Wages of Local Government Employees / Ministry of Internal Affairs and Communications |
| F_cap (Firms) | F_K | Actual value of manufacturing industry tangible fixed assets at the end of the year | Census of Manufactures / Ministry of Economy, Trade and Industry |
| | F_R | Total number of executives and managers dedicated to (air) pollution prevention who have passed national tests | Environmental Management / Japan Environmental Management Association for Industry |
| | F_P | Value of products shipped by manufacturing industry/raw materials value | Annual Report on National Accounts / Cabinet Office |
| C_cap (Citizens) | C_K | Actual expenditures on communications, publications and other printed matter | Annual report of family income and expenditure survey / Statistics Bureau |
| | C_R | Actual city social education expenses | The Annual Report of Local Finances / Ministry of Home Affairs |
| | C_P | Actual expenditures on buses and train passes for commuting to work and school | Base Linked Consumer Price Index Time Series / Statistics Bureau |
| | | | Annual Report on National Accounts / Cabinet Office |
| ENV. (Environment) | SO2 | $SO_2$ concentration (annual average values under ambient air pollution monitoring station) | Air Pollution in Japan / Ministry of Environment |
| | NO2 | $NO_2$ concentration (annual average values under ambient air pollution monitoring station) | |

Source: The Authors

**2.3.2 Firm environmental management capacity (F_cap)**

For firms, we can set resource efficient production (F_P) as the output of air pollution countermeasures. Behind this achievement are the accumulation of technology and expertise related to resource efficient production (F_K) through cleaner production technology, and the organizational structures for pollution prevention and production process management (F_R). Thus, the production process series related to air pollution can be described as F_K to F_R to F_P. In this case, the capacity elements that correspond to these are represented in Table 3 by the 3 items under F_cap. In this table, we applied a GDP deflator to realize actual values of F_K. Moreover, "tangible fixed assets at year-end" (F_K) are not limited to just cleaner production technology, but are assets related to overall production activity. We can use this, though, because we can assume that it is rational behavior for firms to introduce equipment and facilities for resource efficient production and to accumulate technology and expertise for their operation[11]. In addition, according to "total number of executives and managers dedicated to (air) pollution prevention who have passed national tests" (F_R), since pollution prevention managers and others who have passed national tests are tested at 9 different site throughout Japan, we calculated their number as: (total number of people who passed the test at each site)×(applicable urban population / population in jurisdiction of the applicable Regional Bureau of Economy, Trade and Industry) (Honda, 2004). Then, since the average age of the people who passed the first test in 1971 was about 30, we assumed that these personnel and their knowledge and expertise regarding pollution prevention have made continuous contributions, so we treated these years as accumulated data.

Regarding capacity base administration, Konno and Nonaka (1995) have classified that capacity in 3 layers: organizational resources level (knowledge resources, cognitive capacity), knowledge conversion level (systems, processes, skills), and product level (core products, complementary products and services). They state that the interrelations between these result in a firm's total capacity. This corresponds to the items in this research expressed as a firm's K, R and P.

**2.3.3 Citizen environmental management capacity (C_cap)**

We set the active use of public transportation (C_P) as the citizen consumption output of air pollution countermeasures. Behind the realization of this behavior is knowledge and information that cultivate environmental consciousness (C_K) and the creation of situations that promote individual consciousness toward environmental behavior (C_R). The consumption process series related to air pollution can be seen as C_K to C_R to C_P. The capacity elements that correspond to these are shown in Table 3 by the 3 items under C_cap. Moreover, we adjusted C_K and C_P for each city and item using deflators. We used education expenditures as the deflator for C_R.

Stern (2000) has summarized research on individual environmental behavior and the primary factors that influence it and classified environmental behavior as "1. environmental activism," "2. non-activist behaviors in the public sphere," "3. private-sphere environmentalism," and "4. other."

The primary factors of these environmental behaviors could be classified as "(1) attitudinal," "(2) personal capabilities," "(3) external/contextual forces," "(4) habit and routine," and it could be recognized that specific environmental behavior is induced through the mutual interaction of multiple primary factors.

In this analysis, we used environmental public transportation system use as behavior "3. private-sphere environmentalism" (C_P). For its primary factors, we set (1) attitudinal and (2) personal capabilities as "knowledge and information that cultivate environmental consciousness" (C_K) and (3) external/contextual forces and (4) habit and routine as "the creation of situations that promote individual consciousness toward environmental behavior" (C_R).

"Actual expenditures on communications, publications and other printed matter" (C_K) represent thoughts, senses of value and fundamental capacity levels regarding the environment. Next, we looked at personal and mass

Figure 4  Knowledge and Information Related to Air Pollution and Environmental Problems
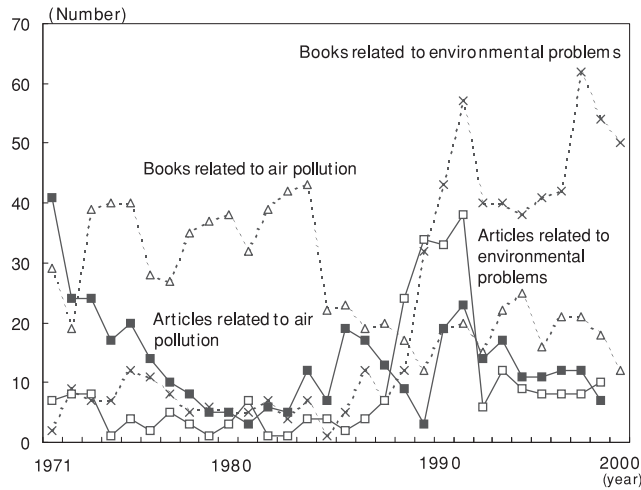
communication classifications, which are information circulation structure classifications related to the creation of Hiromatsu's (1986) information index. Telephones, postcards and similar communication expenses were counted as personal communications, while newspapers, magazines and other publications and printed matter expenses were considered mass communication, and these together formed our information and knowledge index.

Moreover, newspapers and magazines show not only knowledge and information related to basic capacity formation; they can also regularly provide knowledge and information related to air pollution and environmental problems as shown in Figure 4[12].

"Actual city social education expenses" (C_R) includes public meeting place expenses, library expenses, and social education activity expenses and contributes to social standards and the maintenance of customs and manners in the public sphere. In addition, externality of education is also a factor. In short, if decisions about education needs are entrusted to individuals, there is a chance that the desired education level of society as a whole may not be achieved (Oshio, 2002), so education instituted by society complements individual knowledge and information, and encourages activity as a group. Regarding these, knowledge originates amid specific relationships between times, places and other people (Nonaka *et al.*, 2003), so shared education can be said to contribute to the creation of these opportunities. In addition, education by society supports the activities of groups involved in such education, including protection of the natural environmental and as watchdogs for environmental problems.

For this analysis, we set $SO_2$ concentration and $NO_2$ concentration as the annual average values from ambient air pollution monitoring station[13].

## 3. Analysis Results

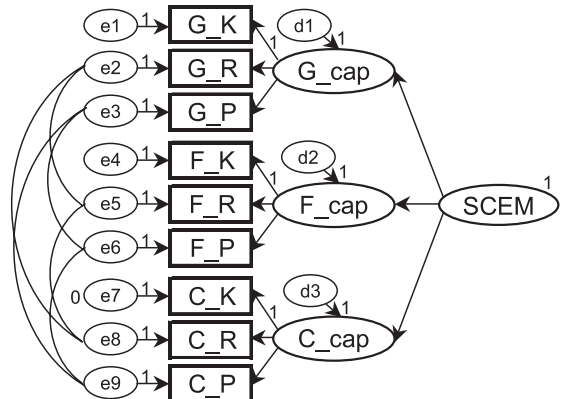### 3.1 Analysis 1: verification of capacity formation
Using the confirmatory factor analysis model in Figure 5, we conducted multi-group simultaneous analysis with results as shown in Table 4.

Figure 5  Primary Factor Analysis Model
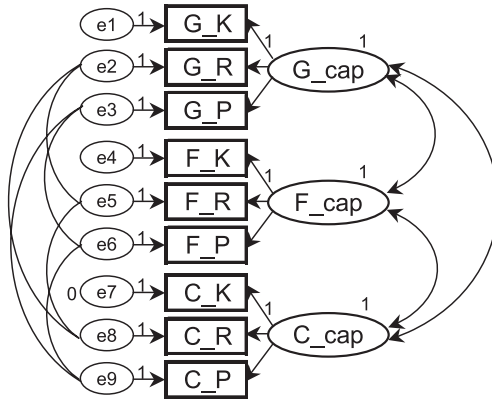


Source: The Authors

Figure 6  The Secondary Factor Analysis Model



Source: The Authors

Since the correlation between each actor's environmental management capacity is high, we hypothesized the existence of shared higher latent concepts as the social capacity for environmental management and created the secondary factor analysis model in Figure 6, with results as shown in Table 5.

In each model, we set the covariance between each actor's R and P error variables[14]. In addition, to assure discrimination for each model[15], for the primary factor model in Figure 5, we fixed each actor's environmental management capacity variance as 1. For the secondary factor model in Figure 6, we sought a solution by setting SCEM variance as 1, and the latent variables to K paths coefficient was set as 1. In addition, judging them to be sample variances (Kano and Miura, 2002; Toyoda, 2003), we set both cities' e7 error variables and Osaka's e4 error

Table 4  Results of Primary Factor Analysis Model

|  |  |  | Osaka | Kitakyushu |
|---|---|---|---|---|
| G_cap | <-> | F_cap | 0.88 *** | 0.76 *** |
| G_cap | <-> | C_cap | 0.94 *** | 0.78 *** |
| F_cap | <-> | C_cap | 0.94 *** | 0.95 *** |
|  | --> | G_K | 0.90 *** | 0.73 *** |
| G_cap | --> | G_R | 0.77 *** | 0.87 *** |
|  | --> | G_P | 0.94 *** | 0.85 *** |
|  | --> | F_K | 1.00 *** | 0.69 *** |
| F_cap | --> | F_R | 0.90 *** | 0.99 *** |
|  | --> | F_P | 0.93 *** | 0.82 *** |
|  | --> | C_K | 1.00 *** | 1.00 *** |
| C_cap | --> | C_R | 0.38 * | 0.72 *** |
|  | --> | C_P | 0.70 *** | 0.29 |
| $\chi 2$=50.399(df 39), P=0.104, GFI=0.807 | | | | |

Note: *** p< 0.01,  ** p< 0.05,  * p< 0.10
Source: The Authors

Table 5  Results of Secondary Factor Analysis Model

|  |  |  | Osaka | Kitakyushu |
|---|---|---|---|---|
| SCEM | --> | G_cap | 0.94 *** | 0.79 *** |
|  | --> | F_cap | 0.93 *** | 0.96 *** |
|  | --> | C_cap | 1.00 *** | 0.99 *** |
|  | --> | G_K | 0.90 - | 0.73 - |
| G_cap | --> | G_R | 0.77 *** | 0.87 *** |
|  | --> | G_P | 0.94 *** | 0.85 *** |
|  | --> | F_K | 1.00 - | 0.69 - |
| F_cap | --> | F_R | 0.90 *** | 0.99 *** |
|  | --> | F_P | 0.93 *** | 0.82 *** |
|  | --> | C_K | 1.00 - | 1.00 - |
| C_cap | --> | C_R | 0.38 * | 0.72 *** |
|  | --> | C_P | 0.70 *** | 0.29 |
| $\chi 2$=50.399(df 39), P=0.104, GFI=0.807 | | | | |

Note: *** p< 0.01,  ** p< 0.05,  * p< 0.10
Since the paths to K are set as1 in order to achieve model discrimination,
CR (Critical Ratio) are not calculated.
Source: The Authors

variable variances as $0$[16].

We calculated model fitness jointly for both populations and found $\chi^2$ =50.399 (df 39), P=0.104 (>0.05), and GFI=0.807 for both models, indicating a degree of fitness[17].

In this $\chi^2$ test, by making the null hypothesis be "the model is correct" since the hypothesis will not be rejected with a P value greater than 0.05, a model with a 5 percent significance level can be adopted. Moreover, GFI is somewhat low, but this is the result of the small sample size (Kano, 1998)[18]. All of the standardized path coefficients are positive, high values, so we can say that the latent variables were suitably measured by the observed variables and the ability to explain cause and effect is high. In addition, except for Kitakyushu's C_cap to C_P, every path coefficient was statistically significant. From this, we can say that structures of social capacity for environmental management (SCEM) for both cities were equally expressed. Moreover, Kitakyushu's C_cap to C_P path that was not statistically significant was reconfirmed in analysis 2.

At this point, interpretation of the model in Figure 6, which is formed from regression models, can be expressed, for example, by the following equation for the top SCEM to G_cap to G_K causal relationship for Osaka.

Figure 7  The Causal Relationship between Capacities and Air Quality (Osaka)
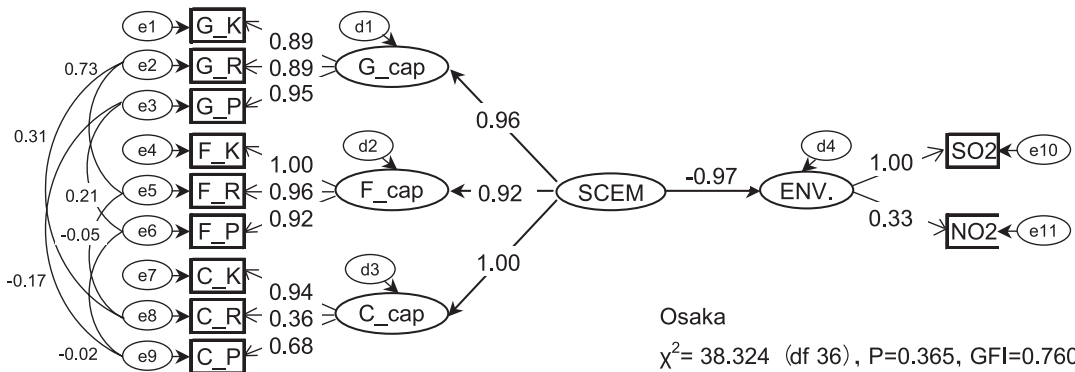


Source: The Authors

Figure 8  The Causal Relationship between Capacities and Air Quality (Kitakyushu)



Source: The Authors

Table 6  Model for each of the Actors

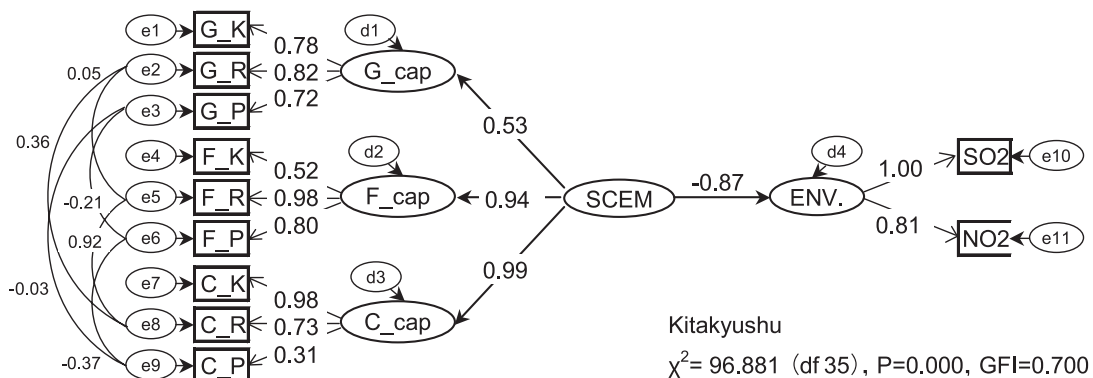| | | | Osaka | | Kitakyushu | |
|---|---|---|---|---|---|---|
| G_cap | --> | G_K | 0.90 | *** | 0.82 | *** |
| | --> | G_R | 0.96 | *** | 0.91 | *** |
| | --> | G_P | 0.72 | *** | 0.82 | *** |
| | --> | ENV. | -0.99 | *** | -0.73 | *** |
| ENV. | --> | SO2 | 1.00 | - | 1.00 | - |
| | --> | NO2 | 0.34 | * | 0.82 | *** |
| individual | | χ2=8.784(df 5) | | | χ2=10.128(df 5) | |
| | | P=0.118 | | | P=0.072 | |
| | | GFI=0.879 | | | GFI=0.860 | |
| simultaneous | | χ2=18.912(df 10), P=0.041, GFI=0.870 | | | | |

| | | | Osaka | | Kitakyushu | |
|---|---|---|---|---|---|---|
| F_cap | --> | F_K | 0.94 | *** | 0.53 | ** |
| | --> | F_R | 0.99 | *** | 0.95 | *** |
| | --> | F_P | 0.98 | *** | 0.84 | *** |
| | --> | ENV. | -0.94 | *** | -0.93 | *** |
| ENV. | --> | SO2 | 1.00 | - | 1.00 | - |
| | --> | NO2 | 0.49 | ** | 0.85 | *** |
| individual | | χ2=11.030(df 5) | | | χ2=10.069(df 5) | |
| | | P=0.051 | | | P=0.073 | |
| | | GFI=0.848 | | | GFI=0.861 | |
| simultaneous | | χ2=21.100(df 10), P=0.020, GFI=0.854 | | | | |

| | | | Osaka | | Kitakyushu | |
|---|---|---|---|---|---|---|
| C_cap | --> | C_K | 0.96 | *** | 1.00 | *** |
| | --> | C_R | 0.74 | *** | 0.72 | *** |
| | --> | C_P | 0.73 | *** | 0.35 | ** |
| | --> | ENV. | -0.98 | *** | -0.84 | *** |
| ENV. | --> | SO2 | 1.00 | - | 1.00 | - |
| | --> | NO2 | 0.33 | * | 0.81 | *** |
| individual | | χ2=8.299(df 5) | | | χ2=8.203(df 5) | |
| | | P=0.141 | | | P=0.145 | |
| | | GFI=0.907 | | | GFI=0.909 | |
| simultaneous | | χ2=16.501(df 10), P=0.086, GFI=0.908 | | | | |

Note: *** p< 0.01,  ** p< 0.05,  * p< 0.10
Source: The Authors

Figure 9  Model for each of the Actors
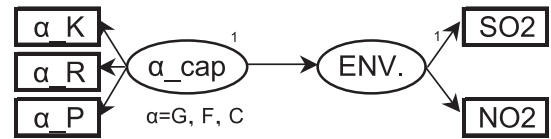


Note: Error variable and disturbance variable are omitted.
Source: The Authors

G_cap=0.94×(SCEM)+d1                (1)

G_K=0.90×(G_cap)+e1                (2) (d1: disturbance variable, e1: error variable)

This shows that the high level of social capacity for environmental management is related to the level of government environmental management capacity being high and also the level of government capacity to provide knowledge, information, and technology (K) being high. Thus, we showed that the level of latent SCEM can be estimated from observed data.

**3.2 Analysis 2: verification of the causal relationship between capacities and air quality**

Using the structural equation model, causal relationship between the social capacity for environmental management estimated using the secondary factor model of analysis 1 and the air quality improvement is as shown in Figure 7 and Figure 8. In order to achieve model discrimination, in each figure the highest path coefficient for the paths from latent variables (path to K, SCEM to G_cap, ENV to SO2) was set as 1 and a solution was sought with this constraint. In addition, we evaluated the influence of sample variance, and set the variance of the error variables e10 for both cities and e4 for Osaka as 0.

The model fitness for Osaka was $\chi^2$= 38.324 (df 36), P=0.365 (>0.05), and GFI=0.760, while for Kitakyushu it was $\chi^2$=96.881 (df 35), P=0.000 (<0.05), and GFI=0.700[19]. In addition, standardized path coefficient values were large. SCEM to ENV was negative, but the rest were all positive, showing that increasing the SCEM and its component elements has a causal relationship with the reduction of air pollution. In addition, all were significant at 5 percent with the exceptions of Osaka's C_cap to C_R and ENV to NO2, and Kitakyushu's C_cap to C_P[20].

The Osaka model shows a certain degree of fitness and the contribution of social capacity for environmental management to air quality improvement was verified.

On the other hand, the fitness of the Kitakyushu model was not especially good. One probable main reason for this is that there were many observed variables in this model and only a small number of samples (Toyoda, 2003).

For this reason, using a simpler model, we sought to confirm the causal relationship between environmental management capacity and air quality improvement for Kitakyushu and verify the paths that were not statistically significant at analyses 1 and 2. For this purpose, we used a structural equation model for each of the actors in Figure 9 to verify the causal relationship of each actor's environmental management capacity and air quality improvement[21]. Moreover, capacity assessment for each actor were conducted on actual assessment sites and confirmation and consideration of causal relationships were conducted not only for the whole but also for the parts. To achieve model discrimination, G_cap (F_cap, C_cap) variance was set as 1 and the ENV to SO2 path coefficient was set as 1 and solutions were sought with these constraints. In addition, we set SO2 error variable variance as 0, judging it to be the influence of sample variance.

Table 6 shows that the fitness of all the models is good and that all the path coefficients are statistically significant. From this, we confirmed each actor's environmental management capacity formation and contributions to air quality improvement, complementing the results of analyses 1 and 2. In addition, through simultaneous analysis, a certain fitness was also achieved for distribution invariance and the causal relationship structures of each actor's environmental management capacity and air quality improvement for both cities can be said to have been shown equally. Thus, we verified the contribution of social capacity for environmental management to air quality improvement.

Based on the above verification from observed data of the capacity formation of firms and citizens, in addition to governments, we have shown the applicability of the capacity assessment framework in Table 1 for an assessment standards system (index system) for estimating the level of SCEM for air quality improvement.


## 4. Conclusion

In this article, by a confirmatory factor analysis and a structural equation model, we estimate the effects of social actors (i.e., government, firms, and citizens) and factors (i.e., policies & measures, resource management, and knowledge & technology) on urban air quality management in 2 major cities (i.e., Osaka, Kitakyushu) in Japan. We showed the following results.

(1) We estimate the structure of SCEM, which contributes to urban air quality improvement, by identifying the actors and factors constituting the social capacity explicitly. Therefore, we propose the Actor-Factor Matrix as a capacity assessment framework that can estimate the level of SCEM from observed data.

(2) We estimate that structures of social capacity for environmental management (SCEM) for both cities were equally expressed by a confirmatory factor analysis.

(3) We verified the contribution of social capacity for environmental management (SCEM) to air quality improvement by a structural equation model. Our results also show the difference of structure in SCEM and contribution to urban air quality by each city.

To solve environmental problems that are becoming more diverse and complex, not only governments, but also firms and citizens need to fulfill appropriate roles. For this to be possible, though, each actor must have attained the capacity necessary to fill its role. Amid a tide of changing political systems, including democratization, devolution,

privatization and private sector activation, effective, efficient solutions for social problems are needed not only for the environment, but also for other issues such as development assistance, government administration, and community planning. These should be solved through the participation and cooperation of diverse actors. Capacity assessments for setting the division of roles for each actor and verification of efficacy are needed along with medium and long-term capacity development plans that take these into account. We believe that the Actor-Factor Matrix can be applied to these types of capacity assessments.

## Notes

1 Civil society includes NGOs, NPOs and similar groups, but in this paper, we focus on empirical analysis citizens as the subject. Furthermore, in our definition of social capacity for environmental management, we use the word "citizen" as representative of the range of civil society.

2 Committee on Japan's Experience in the Battle against Air Pollution (1997) positions $SO_2$ as an air pollutant caused by industrial pollution and $NO_2$ as an air pollutant caused by cities and lifestyles. Polluters and victims are distinct for air pollutants caused by industrial pollution, but the polluters and victims are the same for air pollution caused by cities and lifestyles.

3 The period of analysis for data related to estimating the environmental management capacity of corporations was set to start from 1971 because the Law Concerning the Improvement of Pollution Prevention Systems in Specific Factories, which required the establishment of human organizations with expert knowledge related to pollution prevention in factories, was established in June 1971, and pollution prevention management institutions were started then.

4 Moreover, in terms of the financial scale and the number of staff in the administration, Osaka is the largest, followed by Kyoto and Kitakyushu, the city with the smallest scale. In terms of population, Osaka has the highest population density and Kitakyushu has the lowest.

5 See Osaka city (1994), Kitakyushu City (1998), Katsuhara (2001) and Fujikura (2002).

6 Factor analysis was first presented by Spearman, a scholar from the UK, as a technique to analyze the structure of intelligence by extracting factors that are considered intelligence in test results in classic works on French, English and other languages at the beginning of the 20th century (Yanai et al., 1996).

7 Multi-group simultaneous analysis is a type of analysis that investigates a model of homogeneity between groups. It investigates whether the observed variables measure each factor between groups for configural invariance (Toyoda, 2003). As a further step in the investigation, based on the configural invariance results, equivalent value limitations could be put on particular path coefficients and error distributions, and statistical investigations of the differences among path coefficients is possible (measurement invariance). The number of samples in this analysis is small, however, making the possibility that solutions might not converge high, so we do not conduct analysis with equivalent value limitations.

8 The factor analysis in analysis 1 is also a subordinate model of SEM. In addition, structural equation models can also be called models for conducting factor analysis and regression analysis simultaneously. Moreover, we used Nonaka and Konno's (2003) concept creation methodology, which has steps as follows: 1) observation (idea prototype generation), 2) generalization (mechanism understanding), 3) modeling (causal relationship discovery), 4) application (index and measuring of variables). A structural equation model is used as a method for 3) modeling and 4) application of the concept of social capacity for environmental management.

9 When considering personnel, not only the number of people, but also the quality of their work is important. For empirical analysis of human capital in economic growth, Barro (1991) mainly use the number of years of education, laborer income and other representative variables as measurements of the quality of labor. Considering this, we set the average number of years employed as an indicator of experience and skill to measure the quality of personnel involved in environmental policies, and

multiplied this by the number of staff to determine the overall value of personnel.

10 In fact, according to Hashimoto (1998), when reconsidering the 1978 environmental standards for $NO_2$, only the results of scientific deliberations by a pollution countermeasure commission expert committee of the Central Council for Environmental Pollution Control was consulted, and the revision of environmental standards was determined by administration judgments based on policy decisions, referring to the contents of that report. So, policy decisions were made with clear divisions between the roles of science and policy.

11 Since EP (End of Pipe) technology in pollution countermeasures does not directly raise the efficiency of resource use, CP (Cleaner Production) technology was the subject of analysis as a basic pollution countermeasure. Moreover, there is no data on EP technology for each separate city.

12 Books related to air pollution and environmental problems: the number of NDL-OPAC search results for Japanese books that contain the terms "air pollution" and "environmental problems" in their titles by year of publication. Articles related to air pollution and environmental problems: the number of newspaper articles that contain the terms "air pollution" and "environmental problems" in their headlines found in a database of the Asahi Newspaper morning, evening and Tokyo editions since the Second World War by year.

13 Ordinarily, pollutant concentration distributions in environmental media are expressed by a lognormal distribution and, even a histogram that is close in form to a lognormal distribution. For this reason, in order to use analysis methods that hypothesize a multivariate normal distribution as the population distribution, we made SO2 and NO2 concentrations logarithmic to approximate a normal distribution in our analysis.

14 In cases where there were no obvious reasons to assume otherwise, normally the covariance between error variables is assumed to be 0, but in cases when common variance factors other than latent variables were estimated to exist, we can assume covariance between error variables (Toyoda, 2003). For this reason, the relatively easily transferable explicit knowledge that corresponds to R and P, in comparison with the implicit knowledge that corresponds to K indicated by Nonaka and Takeuchi (1996), has high relative fluidity, so we estimated that primary factors other than for those stipulated by each actor's capacity for environmental management also have an influence and we set covariances.

15 A situation where a model cannot be discriminated is when a single solution for the sought free parameter and the equation's numerical relationship in the assumed model cannot be determined. For this reason, in order to achieve discrimination for the model, it is necessary to apply constraints to, for example, fix the path coefficient and variance. A common constraint is to set latent variable variance as 1, and one path coefficient from the latent variable to the observed variable as 1.

16 Error variance is always nonnegative. However, for an observed variable with very strong factor loads (path coefficients), the error variance becomes a small value and there is the possibility that it might even fall into a negative range. In such cases, error variance can be estimated and fixed at 0 (Toyoda, 2003).

17 In addition, RMSEA is another index for fitness that is frequently used. In this model, RMSEA = 0.071 (< 0.08), showing a certain degree of fitness.

18 The greater the sample number, the more reliable the results, but the necessary minimum sample number depends on the model, and there is no absolute standard (Toyoda, 2003). Toyota (1998) suggests that a sample number of 30 can provide a stable solution in estimation of solutions for the specific concepts of specific individuals.

19 The reason that the degree of freedom (df) differs for both cities is that the distribution of e4 is restricted to 0 only for Osaka. Regarding whether this distribution of e4 should be viewed as a free parameter, Kano and Miura (2002) have stated, "How the degree of freedom of the models should be defined remains a difficult and unresolved problem." Moreover, with the Osaka df of 36, P = 0.365 (> 0.05), which is large, but with a df of even 35, we can achieve significance at 5%.

20 Among paths from latent variable, the highest path coefficients (the path toward K, SCEM to G_cap, and ENV to SO2) in the figure were set at 1 in order to achieve discrimination for the model. CR (Critical Ratio) were not calculated, so significance could not be evaluated, but analysis 1 resulted in judgments of significance.

21 From the primary factor model of analysis, the correlation with each actor's environmental management capacity was strong. Multicollinearity occurs when using the same model for verification of the relationship between the 3 actors' environmental management capacities and air quality improvement, so we verified using a simple model for each actor separately.

# References

Barro, R. J. (1991). Economic Growth in a Cross Section of Countries. *Quarterly Journal of Economics*, 106, 407-443.

Committee on Japan's Experience in the Battle against Air Pollution (1997). *Japan's experience in the battle against air pollution* (in Japanese). Tokyo: The Japan Times.

Fujikura, R. (2002). SOx control measures of local government in Japan. In Terao, T. and Otsuka, K. (Eds.), *Dynamism of the policy process in development and the environment* (in Japanese). Chiba: Institute of Developing Economies, 37-78.

Harashima, Y. and Morita, T. (1995). A comparative Analysis of Environmental Policy Development Process in East Asian countries (in Japanese). *Planning Administration*, 18 (3), 73-85.

Hashimoto, M. (1988). *Environmental Administration* (in Japanese). Tokyo:Asahi Shimbun.

Hiromatsu, T. (1986). Information technology Indicator and Information census. In Hayashi, S. and Nakamura, T. (Eds.), *Japanese economy and economic statistics* (in Japanese). Tokyo: University of Tokyo Press, 97-115.

Honda, N. (2004). The Role of the Social Capacity for Environmental Management in Air Pollution Control: An Application of Three Pollution Problems in Japan (in Japanese). *Papers on Environmental Information Science*, 18, 331-336.

Janicke, M. and Weidner, H. (Eds.) (1997). *National environmental policies: a comparative study of capacity-buildin*. Berlin: Springer.

Japan Society of Atmospheric Environment's Committee for Research on Historical Materials (2000). *History of air pollution of Japan* (in Japanese). Tokyo: LATTICE.

Japan Society of Air Pollution (1993). *Global Air pollution and measures* (in Japanese). Tokyo: Ohmsha.

Kano, Y. (1998). Analysis for data of favorable impression to entertainer. In Toyoda, H. (Eds.), *Covariance structure analysis: a case study* (in Japanese). Kyoto: Kitaouji Shobo, 9-21.

Kano, Y. and Miura, A. (2002). *Graphical modeling and Multivariate statistical analysis of AMOS, EQS and CALIS* (in Japanese). Kyoto: Gendai-Sugakusha.

Katsuhara, T. (2001). *Development and Environmental Problem in East Asia* (in Japanese). Tokyo: Keiso Shobo.

Kitakyushu City (1998). *Environmental pollution control in Kitakyushu city, Environmental pollution control in Kitakyushu city: analysis* (in Japanese). Fukuoka: Kitakyushu City.

Konno, N. and Nonaka, I. (1995). *Intellectualizing capability* (in Japanese). Tokyo: Nihon Keizai Shimbun.

Matsuoka, S. and Kuchiki, A. (Eds.) (2003). *Social capacity development for environmental management in Asia: Japan's environmental cooperation after Johannesburg Summit* (in Japanese). Chiba: Institute of Developing Economies.

Matsuoka, S., Okada, S., Kido, K. and Honda, N. (2004). Development of Social Capacity for Environmental Management and Institutional Change (in Japanese). *Journal of International Development Studies*, 13 (2), 31-50.

Matsushita, K. (2002). *Environmental Governance* (in Japanese). Tokyo: Iwanami Shoten.

Murakami, K. and Matsuoka, S. (2008). Evaluation of Social Capacity for Urban Air Quality Management, *Japanese Journal of Evaluation Studies*, 8 (2), 65-81.

Nonaka, I. and Takeuchi, H. (1996). *The knowledge-creating company: how Japanese companies create the dynamics of innovation* (in Japanese). Tokyo: Toyo Keizai.

Nonaka, I., Izumida, H. and Nagata, A. (2003). *Toward the theory of knowledge-based country: a new paradigm of the policy process* (in Japanese). Tokyo: Toyo Keizai.

Nonaka, I. and Konno, N. (2003). *Methodology of knowledge creation* (in Japanese). Tokyo: Toyo Keizai.

OECD/DAC (1999). Donor Support for Institutional Capacity Development in Environment: Lessons Learned. *Evaluation and Effectiveness*, 3, OECD.

Osaka city (1994). *Environmental pollution control in Osaka city* (in Japanese). Osaka: Global Environment Centre Foundation.

Osaka Prefectural Government (2003). *Total Emission Reduction of automobile Nitrogen Oxides and Particulate Matte in Osaka Prefecture* (in Japanese). Osaka: Osaka Prefectural Government.

Oshio, T. (2002). *The economic analysis of the Japanese education* (in Japanese). Tokyo: Nippon-Hyoron-Sha.

Stern, P. C. (2000). Toward a coherent theory of environmentally significant behavior. *Journal of social issues*, 56 (3), 407-424.

Toyoda, H. (1998). *Covariance structure analysis: a primer* (in Japanese). Tokyo: Asakura Publishing.

Toyoda, H. (2003). *Covariance structure analysis: Q & A* (in Japanese). Tokyo: Asakura Publishing.

Weidner, H. and Janicke, M. (Eds.) (2002). *Capacity Building in National Environmental Policy : A Comparative Study of 17 Countries*, Berlin : Springer.

Yanai, H., Shigemasu, Kazuo, Maekawa, S. and Ichikawa, M. (1990). *Factor analysis* (in Japanese). Tokyo: Asakura Publishing.

【Article】

# Three-dimensional analysis incorporating the Weight of Factors for Evaluation －A Case for the University Course Evaluation－

Atsuko Hoshino, Shunichi Kitahara, Kouji Shingyoji, Kazuhisa Adachi

The Faculty of Social and Information Sciences, Jumonji University

hoshinoa@jumonji-u.ac.jp

Masayasu Watai

The Faculty of Human Life

Jumonji University

Hiromitsu Muta

Graduate School of Decision Science and Technology

Tokyo Institute of Technology

## Abstract

Recently, "student ratings" of courses are becoming popular among Japanese colleges and universities to get more information for Faculty Development（FD）．The purpose of this paper is to propose a new three-dimensional analysis incorporating the weight of factors in student evaluations of university courses. It is useful for the policy of FD to consider the importance of these factors. If the factors that are most important to students is known, course performance can be improved more effectively. Conjoint analysis was used to determine the important criteria. We translate not only the results of conjoint analysis（"utility" and "averaged importance"）but "students ratings" data in a three-dimensional bubble graph. This graph makes it possible to show faculties the priority of factors that must be improved in their courses. Four divisions of the graph, named "continued-factors", "reexamined-factors", "improved-factors" and "reduced-factors" represent the features of the variables that are located in each division.

## Keywords

## 1. Background

Recent years have seen widespread use of student course evaluations at Japanese universities. In the United States, students' evaluation of instruction began in the 1950s, and, accordingly, the literature on the student course evaluation started to appear around that time (Falchikov, 2005). About 40% of the U.S. universities and colleges had reportedly conducted course evaluations in 1960 (Stecklein, 1960). In Japan, the emphasis of the 1991 revision of university accreditation standards on faculty development, self-inspection, and self-evaluation promoted the use of course evaluations (Shimura, 1992). According to a survey by the Ministry of Education, Culture, Sports, Science and

Figure 1 Learning Evaluation Cycle
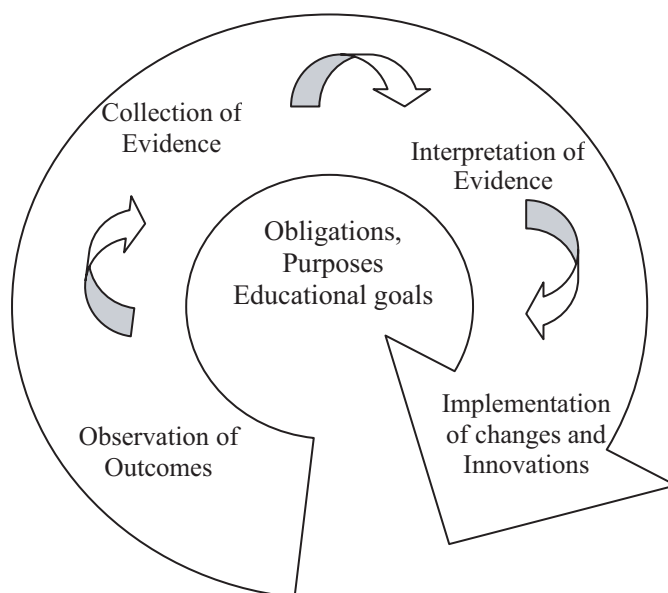


Source: Maki (2004)

Technology (MEXT), merely 334 (47%) of all public and private universities conducted student course evaluations in 1998, although the number has since been increasing. In 2003, for example, course evaluations were used in 96 national universities (99%), 68 other public universities (89%), 469 private universities (99%), and 633 (91%) of all universities.

Braskanp and Ory (1994) ascribe this rise in the use of systematized student course evaluations to (1) a general increase in qualitative evaluations conducted at institutions of higher education, (2) the need for decision-making factors that administrators facing serious financial problems can use, and (3) the capability of efficiently collecting scale-based assessments of classes by students, leading to the quantification of a complicated concept such as the quality of education.

Although student course evaluations are often used for faculty assessment or salary reviews in the United States, currently evaluation results are mainly used for fostering instructional improvements. Maki (2004) proposes a learning evaluation cycle (Figure 1) that starts with the setting of educational goals or purposes that follow certain educational principles. It subsequently leads to the collection of evidence to observe outcomes of instructional and learning activities, then to the interpretation of the evidence, and then to innovations based on the results of the interpretation. After these stages, the same evaluation cycle begins again. This cycle represents the so-called outcomes evaluation with a notable characteristic that evaluations go beyond offering some assessment and lead to improvements.

Course evaluations in Japan generally are not used directly for the assessment of educational performance that is associated with employment issues (Ohyama, 2001). The main purpose of course evaluations is improving instructional techniques in order to raise their educational effectiveness. The MEXT report positions student course evaluations as concrete efforts to improve the quality of courses (University Promotion Division, MEXT Higher Education Bureau, 2005). Now that course evaluations are widely used, many universities are interested in how they

can take advantage of the evaluation results to improve their course instruction.


## 2. Research Purpose

Conducting evaluations should contribute to instructional improvements; however, making improvements to all aspects of a course that receives low ratings is not feasible. Thus, it is necessary to first respond to aspects that are important and are expected to show greater improvements. This paper aims to show empirically a method that presents effective improvement measures three-dimensionally by combining scale-based assessment (rating) and measured importance (weight) of evaluation categories or factors. Though the proposed method is appropriate for the evaluation of policies rather than qualitative evaluations such as course evaluation, the focus of this paper is on course evaluation, a topic which the authors have studied for several years.

There are mainly two measures to assign weights. One is the direct assignment of importance by the evaluator. This includes a grading method in which the evaluators score the importance of each evaluation factor on a scale of one to ten, and a ranking method in which they rank the factors or assign ordinal values reflecting the ranking. The other measure determines the importance of each factor based on a questionnaire regarding preferences of the evaluators. This includes the Analysis Hierarchy Process (AHP) and conjoint analysis. Since students are the evaluators, valid data cannot be expected from the direct assignment of the importance of each factor. Also, the AHP is not appropriate as an evaluation method intended to be administered using part of the class time because it involves repeated comparisons. Hence, this paper uses conjoint analysis in weighting evaluation categories.

Conjoint analysis determines preferred attributes and the importance of each factor using data on the evaluators' preferences associated with the levels and attributes of multiple evaluation factors. It was initially developed by Luce and Tukey (1964) in the field of psychometrics and subsequently became widely used in marketing research, in which Green and Wind (1973) were the first to present a decision model with multiple attributes. The analysis was also incorporated in environmental economics in the 1990s (Kuriyama, 2000). Moreover, it has recently been applied in education and welfare studies though the number is still small (See Maki, 1998; Nagai, 2002, for example). Setting levels for each factor appropriately is most crucial when conjoint analysis is used for course evaluations. Since students can only decide on a level of each factor based on the class they actually have taken, it is necessary to set the levels such that they reflect the characteristics of the class and to clarify in the questionnaire what each level represents.


## 3. Analytical Method

Pre-surveys were conducted before the main survey in four classes (two lecture courses and two applied courses) at three different universities because no data existed on student course evaluations at universities for which conjoint analysis can be applied. The result showed the effectiveness of conjoint analysis in examining student needs for classes. The analysis, which takes into account differences in students' academic record and major, revealed the students did not regard an "easy" course highly, and wanted classes in which students can develop high-performance skills (Hoshino and Muta, 2003; Hoshino, 2004). For the main survey, both conjoint analysis based on profiles and scale-based course assessments were conducted.

The analysis is focused on the computer literacy courses (applied courses) offered in the second semester of the 2002 academic year at an all-female university. A computer literacy course is mandatory for freshmen. There are four

Atsuko Hoshino, Shunichi Kitahara, Kouji Shingyoji, Kazuhisa Adachi
Masayasu Watai   Hiromitsu Muta

Table 1  Factors and Levels Used for the Conjoint Analysis

| Factor / Level | Amount of knowledge transferred | Difficulty | Instructional medium | Assignments | Textbook | Individual instruction |
|---|---|---|---|---|---|---|
| 1 | Small | Low | Mainly blackboard | No | No | No |
| 2 | Large | High | Mainly presentation | Yes | Yes | Yes |

Source: Authors

courses in total: three different courses for three different levels, plus a course for international students. The analysis considers the former three courses. For the rating questionnaire, the size of the sample that is valid is 37, 41, and 38 for the advanced, intermediate, and introductory courses, respectively; it is 40, 44, and 42 for the conjoint analysis. In the conjoint analysis, evaluation factors and their levels are set. Table 1 shows the six factors and two levels used for this study.

A full-profile method, in which multiple factors are simultaneously presented, is used in the survey. Regarding the levels, the survey includes the following questions:  (1) regarding the amount of knowledge transferred, whether the amount of knowledge that just covers the content of the syllabus is sufficient or whether the maximum possible amount of knowledge is preferred;  (2) regarding difficulty, whether covering basic topics is sufficient or whether some advanced topics should be included;  (3) regarding the instructional medium, whether lectures based on the use of the blackboard are preferred or whether lectures with pre-structured materials (e.g., presentation materials) are preferred;  (4) regarding assignments, whether or not assignments (e.g., small essays or small in-class quizzes) should be given;  (5) regarding textbooks, whether or not textbooks should be used; and  (6) regarding individual instruction, whether or not the instructor should move around the classroom providing students with individual instruction when necessary.

Eight profiles are extracted using an orthogonal design with six factors and two levels. Also, two additional hold-out cards are included in order to verify if the model is appropriate. Thus, the survey is based on ten profiles. The hold-out cards are not taken into account in the estimation of Utility. Instead, the rating values for the hold-out cards are computed based on the estimated Utility and are compared with the actual rating values from the respondents. A high correlation between the computed and actual values would indicate the reliability of the estimation results. The questionnaire format is of full profile rating, and all profiles are evaluated on a 5-point scale of "would not want to take the course under the given conditions (= 1)" to "would very much want to take the course under the given conditions (= 5)". Since the levels used for the factors are all in a nominal scale, a discrete model is applied. The method of least squares is used for the estimation of Utility.

As for the rating questionnaire, a total of 20 categories are used. They consist of 15 categories for evaluating the course plus the degree of students' expectation and effort for the class as well as the degree of understanding, attendance rate, and overall satisfaction level. The attendance rate and satisfaction level are measured between 10% and 100%, and a 5-point evaluation is used for the other categories. The importance of the factors are related with the ratings on the evaluation categories by linking the six factors for the conjoint analysis with course evaluation categories and by showing on a bubble graph the Utility and relative importance from the conjoint analysis and course evaluation data for the relevant categories.

## 4. Results and Interpretations

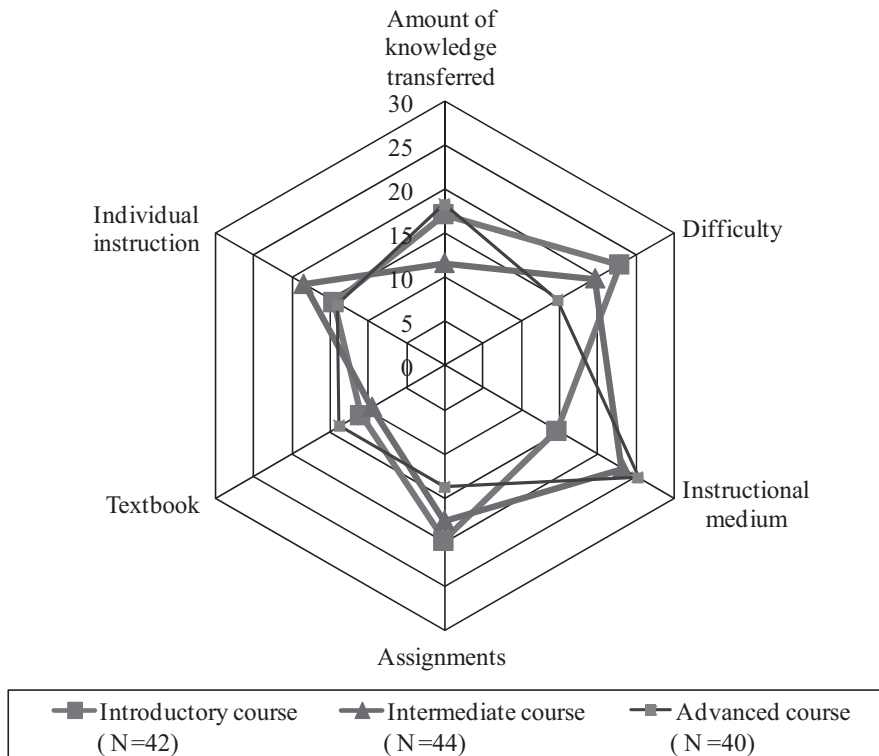### 4.1 Utility and Relative Importance for the Factors

Let us first examine the results of the conjoint analysis. Table 2 shows the Utility and relative importance for each course, and Figure 2 is a graph of relative importance for each course. The students prefer levels with a positive Utility, and the higher the value, the greater the importance of the relevant factor. Relevant importance shows the proportion of a Utility to the whole. Also, the Utility and relative importance in Table 2 are average values because

Table 2  Utility and  Relative Importance

| Factor | Level | Introductory course (N=42) | | Intermediate course (N=44) | | Advanced course (N=40) | |
|---|---|---|---|---|---|---|---|
| | | Utility | Relative importance | Utility | Relative importance | Utility | Relative importance |
| Amount of knowledge transferred | Small | -0.039 | 17.16 | -0.003 | 11.60 | -0.084 | 18.37 |
| | Large | 0.039 | | 0.003 | | 0.084 | |
| Difficulty | Low | 0.313 | 22.95 | 0.264 | 19.70 | 0.022 | 14.83 |
| | High | -0.313 | | -0.264 | | -0.022 | |
| Instructional medium | Mainly blackboard | -0.176 | 14.69 | -0.338 | 23.25 | -0.359 | 25.39 |
| | Mainly presentation | 0.176 | | 0.338 | | 0.359 | |
| Assignments | No, | 0.182 | 19.75 | 0.122 | 17.58 | 0.122 | 13.77 |
| | Yes | -0.182 | | -0.122 | | -0.122 | |
| Textbook | No, | 0.051 | 11.01 | -0.037 | 9.45 | 0.016 | 13.69 |
| | Yes | -0.051 | | 0.037 | | -0.016 | |
| Individual instruction | No, | -0.158 | 14.44 | -0.287 | 18.42 | -0.116 | 13.96 |
| | Yes | 0.158 | | 0.287 | | 0.116 | |
| CONSTANT | | 2.878 | | 2.889 | | 2.966 | |
| Pearson's R | | 0.993 | | 0.999 | | 0.996 | |

Source: Authors

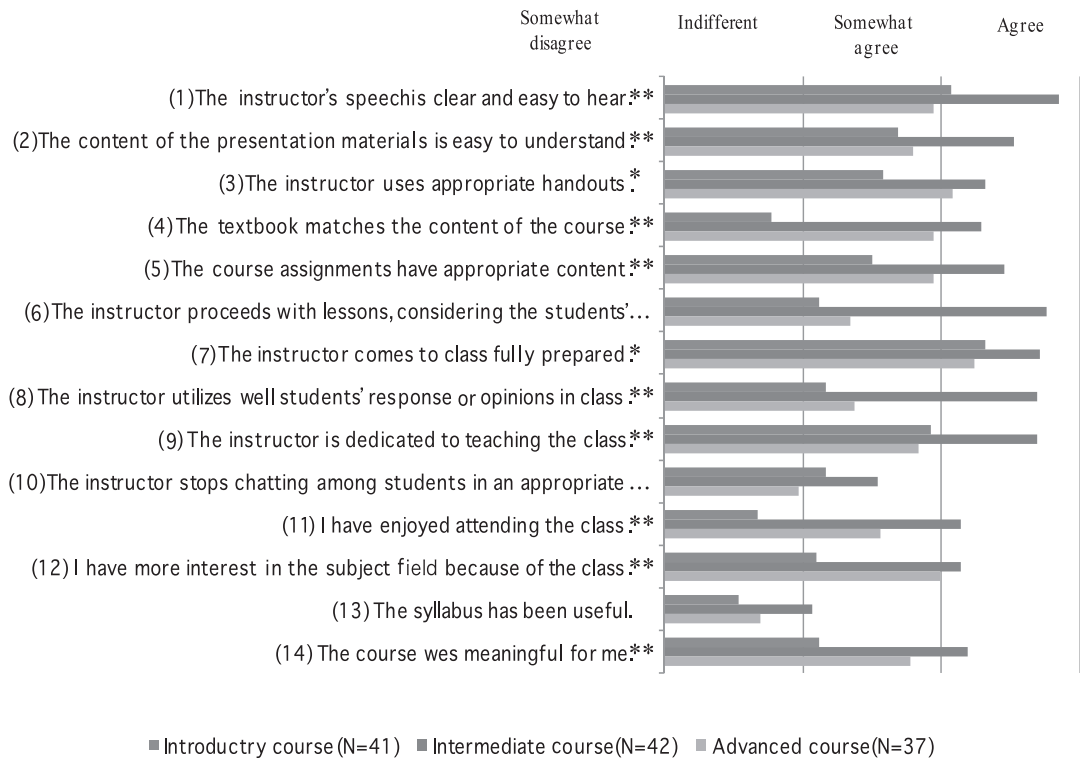Figure 2　Relative Importance of the Factors for the Three Courses



Source: the Authors

the estimated values for the individuals are computed with the use of full profile rating. Therefore, the proportion of the Utility does not necessarily coincide with the associated relative importance in Table 2.

　　Table 2 reveals that factors deemed important vary for different classes: the highest relative importance is observed in the "difficulty" factor for the introductory course and in the "instructional medium" factor for the intermediate and advanced courses. For all three courses the preferred level for each factor is "small" for "amount of knowledge transferred", "low" for "difficulty", "mainly presentation style" for "instructional medium", "no" for "assignments", and "yes" for "individual instructions". As for the "textbook" factor, the preferred level is "yes" for the intermediate course and "no" for the introductory and advanced courses. The factor "amount of knowledge transferred" has a small Utility compared to average relative importance for all classes (0.039 to 17.16% for the introductory course, 0.003 to 11.60% for the intermediate course, and 0.084 to 18.37% for the advanced course). The reason can be a large variance which means that students' preferences are split between "large" and "small". As for "textbook", both the Utility and relative importance are comparatively low. This means that whether or not a textbook is used does not have much impact, which implies that the textbook itself should be reexamined.

　　As for the "difficulty" factor, both the Utility and relative importance are high for the introductory and intermediate courses, indicating the importance of the class being easy to follow. Also, for the advanced course, both the Utility and relative importance are comparatively low at 0.0219 and 14.8%, respectively. This indicates that the difficulty of the course is not so important for advanced-level students, and that students' preferences are split between "low" difficulty and "high" difficulty. As is clear from Figure 2, the lower the course level, the higher the

Figure 3  Average Values for Course Evaluation Categories for the Three Courses

| | Somewhat disagree | Indifferent | Somewhat agree | Agree |
|---|---|---|---|---|

(1)The  instructor's speech is clear and easy to hear.**
(2)The content of the presentation materials is easy to understand.**
(3)The instructor uses appropriate handouts.*
(4)The  textbook matches the content of the course.**
(5)The course assignments have appropriate content.**
(6)The instructor proceeds with lessons, considering the students'…
(7)The instructor comes to class fully prepared.*
(8)The instructor utilizes  well students' response or opinions in class.**
(9)The instructor is dedicated to teaching the class.**
(10)The instructor stops chatting among students in an appropriate …
(11)I have enjoyed attending the class.**
(12)I have more interest in the subject field because of the class.**
(13)The syllabus has been useful.
(14)The course wes meaningful for me.**

■Introductry course (N=41)   ■Intermediate course(N=42)   ■Advanced course(N=37)

Note: ** significant at the 1% level; * significant at the 5% level.

relative importance of the difficulty of the course, that is, a class that is more understandable is desired. As for "individual instructions", relative importance is especially high in the intermediate course. This could be because students' skill levels are widely spread in the intermediate course relative to the other courses and the need to respond to this gap is high.

### 4.2 Overview of the Ratings

A total of 20 evaluation categories are used, which consist of 15 categories for evaluating the course plus the degrees of students' expectation and effort for the class as well as the degree of understanding, attendance rate, and overall satisfaction level. Figure 3 shows the course average values and result of analysis of variance for 14 of the evaluation categories, excluding categories which have many "not applicable" responses. All categories in Figure 3, except "usefulness of the syllabus", are statistically significant for all courses. Generally, the intermediate course gives high ratings and the introductory course low ratings. Focusing on separate categories, one finds high ratings for "speech", "preparation", and "dedication", which reflect efforts by the instructor. Categories associated with communications between the instructor and students, such as "consideration to the students' understanding", "utilization of students' response", and "stopping chatting" show large differences among the courses. Also, differences are seen for categories related to instructional materials such as "presentation materials", "textbook", and "assignments".

Table 3  Factors in the Conjoint Analysis and Course Evaluation Categories

| Conjoint Analysis | | Student Course Evaluation (Rating) |
|---|---|---|
| Factors | Levels | Evaluation Categories |
| Amount of knowledge transferred | Small/ Large | I have more interest in the subject field because of the class. |
| Difficulty | Low/ High | The instructor proceeds with lessons, considering the students' understanding. |
| Instructional medium | Mainlyblackboard,/ Mainly presentation | The content of the presentation materials is easy to understand. |
| Assignments | No/ Yes | The course assignments have appropriate content. |
| Textbook | No/ Yes | The textbook matches the content of the course. |
| Individual instruction | No/ Yes | The instructor utilizes well students' response or opinions in class. |

Source: the Authors

## 4.3 Establishment of Related Evaluation Categories

To conduct an analysis of course evaluation data that is linked with the conjoint analysis, related categories from both are examined. Table 3 shows the related categories. The linkage between the factors in the conjoint analysis and the corresponding evaluation categories is based on the following considerations.

- ・Since the "assignments" and "textbook" factors both have two levels of "yes" and "no", the categories of "the course assignments have appropriate conten" and "the textbook matches the content of the course" are chosen as corresponding evaluation categories.
- ・Since the "instructional medium" has two levels, "mainly blackboard" and "mainly presentation style", and the latter is preferred in all three classes, the category of "the content of the presentation materials is easy to understand" is the corresponding category.
- ・The "amount of knowledge transferred" factor has two levels, "small (only essential knowledge)" and "large (as much knowledge as possible)". Since the degree of interest in the field is considered to have a strong relationship with the amount of knowledge, the category of "I have more interest in the subject field because of the class" is the corresponding category.
- ・Since the "difficulty" factor has two levels, "low (basic course content)" and "high (somewhat advanced content)", and has a strong tie with the degree of students' understanding, the category "the instructor

proceeds with lessons, considering the students' understanding" is the corresponding category.
・As the "individual instructions" factor relates to the response to individual students, the category "the instructor utilizes well students' response or opinions in class" is the corresponding category.

**4.4 Weights to Ratings**

We will now examine the relationship between course evaluations and the factors' Utility and relative importance when the factors of the conjoint analysis are linked with course evaluation categories and the weights of the factors with ratings in course evaluations. To clarify the mutual relationships among the three concepts, we use a bubble graph, which facilitates three-dimensional visualization. Figures 4 through 6 show such graphs for the three courses. The Utility is plotted on the x-axis and course evaluation on the y-axis. The size of a bubble signifies the relative importance of each factor. Shaded bubbles are those with relative importance of 15% or greater.

In addition, in order to clarify the characteristics of the factors with the position as well as the size of the bubbles, borders are created for the Utility (x-axis) and rating (y-axis) and the space is divided into quadrants. The rating "3" corresponds with the response "indifferent". However, the border is set at 3.5 because students tend to give somewhat higher ratings in general. As for the Utility, the border is set at 0.1 since half of the six factors have Utility of 0.1 or greater in most cases in the pre-surveys. Factors in each quadrant are characterized as follows.

1. *Factors to be continued.* Factors with a high rating in the course evaluation and a high Utility are considered to highly satisfy students without any modification. Thus, the factors in this quadrant are labeled as "factors to be continued". That is, these factors should continue to be used in the classroom.
2. *Factors to be improved.* Factors with a high Utility and low rating are those that greatly influence students' overall satisfaction with the course, but receive a low rating. They require improvements and thus are labeled as "factors to be improved".
3. *Factors to be reexamined.* The combination of a low Utility and high rating means that despite the instructor's efforts they have a small impact on students' overall satisfaction with the course. Thus, the role of such factors should be reexamined, and, thus, they are labeled as "factors to be reexamined".
4. *Factors to be reduced.* Factors with a low Utility and low rating mean that the instructor's effort to improve them is likely to have a small impact on students' satisfaction ultimately. Such factors should be eliminated if possible. Otherwise, they should be modified. Thus, the factors in this quadrant are labeled as "factors to be reduced".

Attention should be paid to the relative importance of the factors to be reexamined and to be reduced. A low Utility combined with high relative importance means that the variance of the Utility is large, and that differences among individual preferences are great. Put another way, since both the Utility and relative importance are group average values, the greater the differences, the closer to zero the Utility, which in turn reduces its absolute value. Also, as to relative importance, a high average value means that many students deem the relevant factor important. Hence, high relative importance with a low Utility indicates that the students consider the relevant factor important, and that their preferences vary.

In this regard, it is rather appropriate to consider factors in the "to be reduced" category that have high relative importance as "factors to be improved". However, since the preference level is not clear, the direction of improvements to be made is obscure. Thus, it is not an easy task to set improvement guidelines. Also, factors in the "to be reexamined" category that have high relative importance signify that the instructor's efforts are fruitful, but that students' preferences vary. Hence, they can be regarded as "factors to be continued" because they are highly

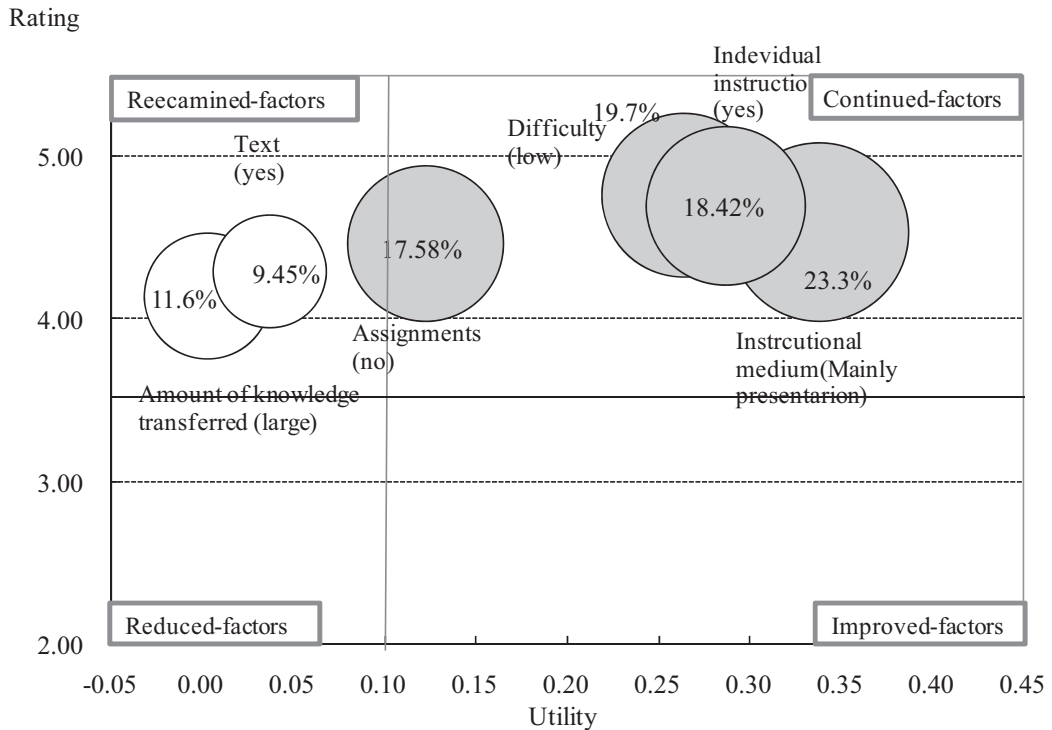Figure 4  Ratings in the Course Evaluation and Utility with Factor Weights (Introductory Course)
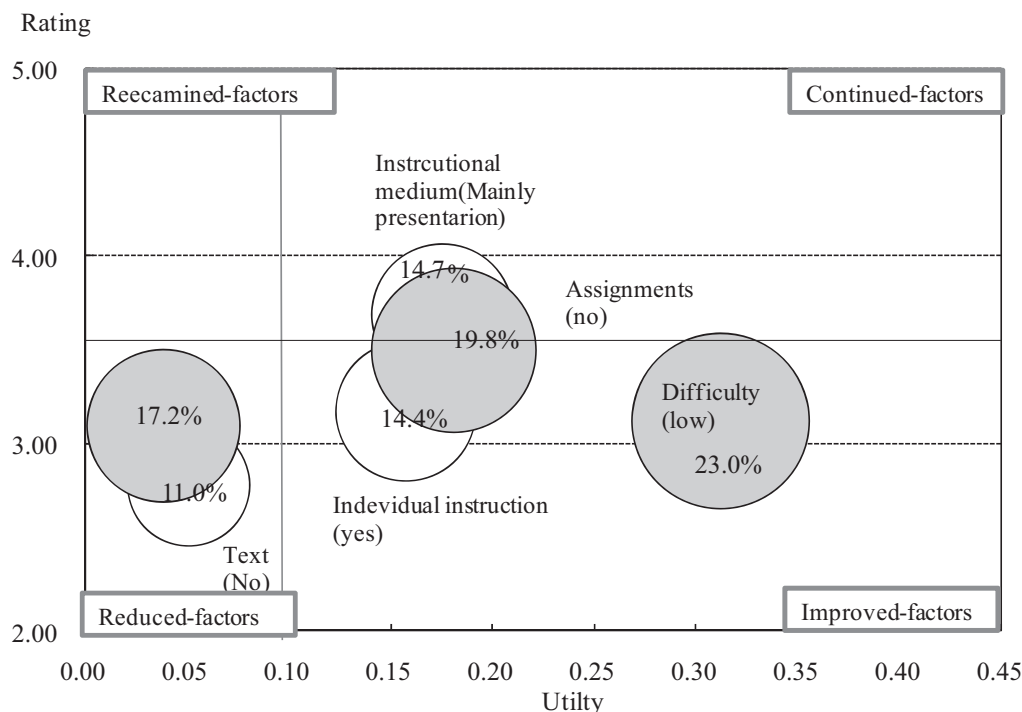
Rating

evaluated despite the dispersion of preferences.

　　Figure 4 shows the bubble graph of the ratings from the course evaluation and Utility for the introductory course. Since the "difficulty" and "individual instructions" factors are categorized as "factors to be improved", modification should be made so that "the instructor proceeds with lessons considering the students' understanding" and "utilizes students' response or opinions in class". The "instructional medium" factor is a borderline case, but the instructor probably should maintain presentation-based instruction and modify the teaching content slightly. Assignments are currently provided, but students prefer not having them. However, since the "assignments" factor sits on the borderline between "factors to be continued" and "factors to be improved", it seems appropriate that the instructor continues to provide assignments, but improve their content. The "amount of knowledge transferred" and "textbook" are categorized as "factors to be reduced". As to the "amount of knowledge" factor, since relative importance is somewhat high, it is desirable to make improvements, responding to the variety in students' preferences. Also, the use of the textbook does not seem to be necessary.

　　Figure 5 shows the bubble graph of the ratings from the course evaluation and Utility for the intermediate course. Since the "instructional medium", "individual instructions", and "difficulty" factors are categorized as "factors to be continued", the current instructional method seems to be effective and should be maintained. Though students prefer not having assignments, they are part of" factors to be continued" and should continue to exist. Both the "textbook" and "amount of knowledge" factors are categorized as "factors to be reexamined". The use of a textbook itself needs to be reviewed, and a determination should be made whether another textbook is used or whether the absence of a textbook is continued. As to the "amount of knowledge transferred", it is hard to imagine that changing it will impact the overall satisfaction level. Thus, no change is needed.

Figure 5  Ratings in the Course Evaluation and Utility with Factor Weights (Intermediate Course)
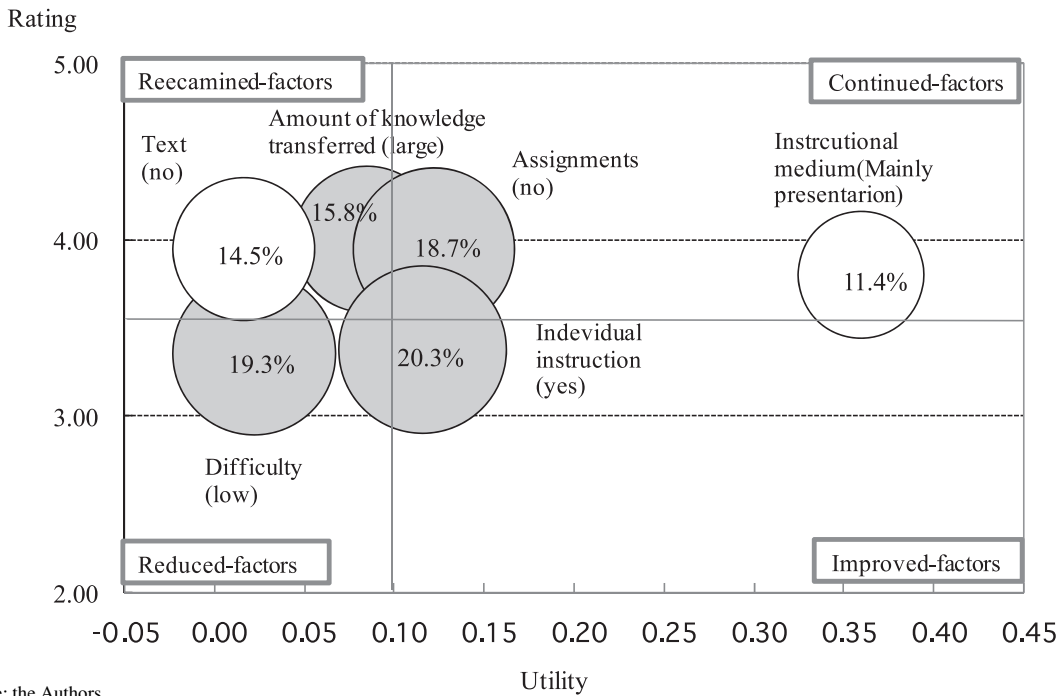


Source: the Authors

Figure 6 shows the bubble graph of the ratings from the course evaluation and Utility for the advanced course. Since the "instructional medium", and "assignments" factors are categorized as "factors to be continued", the current situation should be maintained. The "textbook" factor belongs to the "factors to be reexamined" category, and, thus, its use should be reviewed, as it is the case for the intermediate course. Although the "amount of knowledge" factor is also in the same category, the continuation of the current situation seems appropriate because it is close to the "factors to be continued"  category and has high relative importance.

The "individual instruction" factor is in the "factors to be improved" category. Since it is close to the "factors to be reduced" category, differences in students' preferences are considered to be wide. Thus, the instructor should make great efforts to "utilize students' response or opinions in class well". Although "difficulty" belongs to the "factors to be reduced" category, it can also be regarded as a factor to be improved. Given a wide variability in students' preferences, the instructor should teach students assuming that they have various levels of understanding.

### 4.5 Examination of Student Satisfaction and Total Utility
In student course evaluations, the level of students' satisfaction with the course is critical for the determination of its overall benefits. Hence, the survey result on the satisfaction level is compared with the result of the conjoint analysis. The overall satisfaction level in the conjoint analysis is obtained with estimated total utility for given levels of the factors (i.e., the sum of Utility plus the constant term). The given levels correspond with the characteristics of the actual class. Figure 7 shows the satisfaction level and the estimated values of total utility from the conjoint analysis for the three courses.

Figure 6  Ratings in the Course Evaluation and Utility with Factor Weights (Advanced Course)

Rating

Figure 7  Level of Satisfaction with the Course and Total Utility
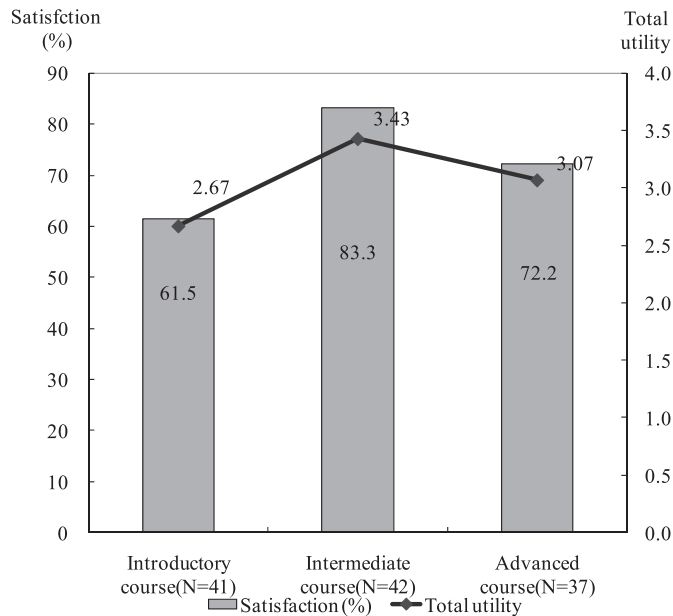


Note: The levels representing the characteristics of the actual course are used: "instructional medium"= "mainly presentation style", "assignments"="yes", "textbook"="yes", and "individual instructions"="yes" (only for the intermediate course).  A value of zero is used for "difficulty" and "amount of knowledge transferred".

The factor levels that are common for all three courses are "instructional medium (presentation)", "assignments (yes)" and "textbook (yes)". Since standards used to determine the "amount of knowledge transferred" and "difficulty" are not obvious, their Utility are set to zero. As for "individual instructions", "yes" is used for the intermediate course and "no" for the introductory and advanced courses. As Figure 7 reveals, the satisfaction level from the course evaluation questionnaire and the estimated total utility values in the conjoint analysis behave in a similar manner. The satisfaction level is 61.5%, 83.3%, and 72.2% for the introductory, intermediate, and advanced courses, respectively. The result of analysis of variance shows that they are significant at the 1% level. Also, the estimated total utility values are 2.67, 3.43, and 3.07, respectively.

The satisfaction level is obtained from the questionnaire in which students were asked to answer how much they were satisfied with their class in percentage terms. The average satisfaction level for each course is similar to its corresponding total utility computed with the factors reflecting the actual condition of the course. Thus, it is considered that students judge their satisfaction level intuitively as well as comprehensively, and that the analytical results are reliable. Hoshino et al. (2005) discuss a detailed analysis of the relationship between the satisfaction level considered in this paper and other evaluation components.


## 5. Summary and Issues for Future Considerations

This paper applies weights obtained through a conjoint analysis to evaluation categories in traditional course rating and examines a method that can utilize course evaluation results for making instructional improvements. Bubble graph analysis enables a visual understanding of the characteristics of the factors (evaluation categories) by three-dimensionally showing the rating, Utility, and relative importance. Also, the four quadrants in the bubble graph facilitate the categorization of factors, based on their characteristics, into "factors to be continued", "factors to be improved", "factors to be reexamined", or "factors to be reduced". The characteristics of factors can be quickly identified from their position and the size of their bubble.

The course evaluation data considered in this paper are not originally intended for the use with a conjoint analysis. Therefore, the evaluation categories linked to a factor are not necessarily matched appropriately. Future tasks include analyses based on the design of factors and evaluation categories, which will lead to instructional improvements, and the use of more objective levels for the factors considered. In addition, since estimated values can be computed for individuals under full-profile rating, further analyses on the effect of other factors such as prior knowledge, academic performance, and attendance rate can be conducted if course evaluation data that identify individuals become available.

The method of analysis proposed in this paper is considered to be appropriate for the evaluation of policies rather than qualitative evaluations such as course evaluations. Based on the result of this paper, we hope to make the three-dimensional analysis of evaluations more practical by applying the analysis to other fields.

## References

Braskanp, A. and Ory, J. C. (1994). *Assessing Faculty Work*. San Francisco, CA: The Jossey-Bass.

Falchicov, N. (2005) *Improving Assessment through Student Involvement*. Abingdon, Oxon: Routledge Falmer.

Green, P.E. and Wind, Y. (1973). *Multiattribute Decisions in Marketing: A Measurement Approach*. Hindsdale, IL: Dryden Press.

Hoshino, A. (2004). A Simulation of University Students' Satisfaction with Courses. *Abstracts of Papers Presented at the Japan Evaluation Society 1st Spring National Conference*, 175-180.

Hoshino, A., Kitahara, S., Shingyouchi, K., Adachi, K., and Watai, M .(2005). Analysis of Influential Factors on Computer Literacy Course Evaluation. *SITE 16th International Conference,* 872-877.

Hoshino, A. and Muta, H. (2003). Instructional Improvements at Universities Based on Conjoint Analysis. *Papers Presented at the Japan Society for Educational Technology 19th Conference*, 691-692.

Kuriyama, K. (2000). Conjoint Analysis, In Ono, E. (Ed.), *Economic Environmental Evaluations in Practice.* Tokyo: Keisoshobo, 105-122.

Luce, R.D. and Tukey J.W. (1964). Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement, *Journal of Mathematical Psychology*, 1, 1-27.

Maki, P.L. (2004). *Assessing for Learning.* VA: Stylus Publishing, LLC.

Nagai, K. (2002). Conjoint Analysis on English Learning Needs of Students at Technical Colleges. *Journal of Tsuyama National College of Technology*, 44, 123-126.

Ohyama, Y. (2001). Issues and Prospects of University Education. *Kyoto University Higher Education Research*, 7, 37-55.

Shimura, E. (1992). University Education and Course Evaluations: The View of the Council for University Establishment and School Incorporation. *IDE Today's Higher Education*, 332, 14-17.

Shinjo, T. (1998). Investigation on Requests by Special Education Teachers in Regarding the Acquisition of the Teaching License in Osaka. *Japanese Journal of Developmental Disabilities*, 20(3), 91-97.

Stecklein, J. E. (1960). Colleges and Universities Programs: Evaluation. In Harris, C.S.(Ed.). *Encyclopedia of Educational Research.* : Macmillan, 285-289.

University Promotion Division, MEXT Higher Education Bureau (2005). *On the Current Situation of Educational Reforms at Universities*, Vol. 2, Ministry of Education, Culture, Sports, Science and Technology.

【Article】

# A Consideration of NGO Accountability in Japan – Accountability on Vision and Strategy, and a Conceptual Classification of NGO Accountability –

## Satoshi Morita

NPO Kokkyo naki Kodomotachi

morita_s@knk.or.jp

## Abstract

While civic organizations have recently been becoming more influential than ever in Japan, NGOs are expected to be accountable for the meaningfulness and effects of their development assistance projects in order to gain understanding from stakeholders. However, there have not been enough discussions on NGO accountability in Japan. Also, a question is raised in such discussions －Who are "NGOs"? This article considers NGO accountability from a holistic viewpoint by applying the case study of the evaluation of a project conducted by an NGO in Japan, based on the assumption that roles and functions of evaluation are important in NGO accountability. This article also considers conceptual classification of NGO accountability. It is concluded that, since the contents of accountability can be different depending on the relative relationship with target stakeholders, it can be useful to: 1) implement several different types of evaluations even in a single project; and 2) involve stakeholders in the evaluation process. Future prospects and issues to deal with: This article also points out: 1) the necessity of building partnerships between NGOs and different actors including the government (or administration), private enterprises and citizens; and 2) the importance of accountability for NGOs' visions and strategies.

## Keywords

NGO, Accountability, Vision and Strategy, Partnership, Mutual Model

## Introduction

In recent years, the public's interest in global-scale issues has been rapidly increasing in Japan. In the meantime, civic organizations[1], including non-government organizations (NGOs) and nonprofit organizations (NPOs), have been more active and thus, they are gaining more presence and influence in society. In December 1998 the Law called *Tokuteihieirikatsudo Sokushinho* (promote nonprofit activities), known as the NPO Law, came into effect. Since then, the number of newly registered NPOs has been increasing. As of the end of December 2008, 33,963 NPOs were registered and certified (the Cabinet Office 2008). According to *NGO Data Book* and *the Directory of International Cooperation NGOs*, published periodically by the Japan NGO Center for International Cooperation (JANIC), the

number of civic organizations in international cooperation (hereinafter referred to as "NGOs[2]") with a total annual revenue exceeding 100 million yen (equivalent to approximately one million US dollar) is 31 in 1992, 42 in 2004, and 46 in 2005. That is, there are quite a few organizations with increasing fiscal scales (Ito, 2000; JANIC, 2007, 2008).

The main financial sources of NGOs include: 1) membership fees; 2) donations from individuals or private organizations; 3) profits from operating funds; 4) revenue from sales of products through events and fair trading implemented independently by the organization; 5) funds from private foundations and governmental institutions; and 6) revenue from government institutions or the United Nations agencies for consigned projects. Out of these sources, 1) to 4) are generally called self-financial sources. One of the recent tendencies among Japanese NGOs is smaller percentages of revenue from self-financial sources. JANIC's data show that revenue from self-financial sources of NGOs is decreasing year by year: 69.3 percent in Fiscal Year 1996, 59 percent in FY2002, and 50 percent in FY2004 (Ito, 2007). On the other hand, the revenue from the government is growing from 3 percent in FY1992 to 10.9 percent in FY2002 (Ito, 2000, 2007). In 2005, the revenue of NGOs in Japan from public institutions reached 12.9 percent[3]. These data imply that: 1) NGOs are more and more dependent on public financial sources for costs of overseas operations; and 2) that revenues from self-financial sources tend to be spent more on management expenses rather than for activities in the field.

Such situations bring a focus on recent partnerships between NGOs and the government in Japan. The Ministry of Foreign Affairs introduced funding systems, such as *NGO Jigyohojokin* (the Grant Assistance for Japanese NGO Projects) in 1989 and *Nihon NGO Shien Mushosikinkyoryoku* (the Grant Aid for Japanese NGO Projects) in 2002. Japan International Cooperation Agency (JICA) started to consign projects to Japanese NGOs in 1999 and have provided assistance through the system of *Kusanone Gijutsukyoryoku Jigyo* (Grass-roots Technical Cooperation) since 2002. In addition, the then Ministry of Posts and Telecommunications introduced the system of *Kokusai Volunteer Chokin* (the Postal Savings for International Voluntary Aid) in 1991 for allocation of funds for NGOs (Shigeta, 2005). While the government's funding for domestic NGOs has been institutionalized and developed as seen above, it is also evident that a number of large-scale NGOs tend to rely on funding from the government as it is not easy for NGOs to ensure self-financial sources for various reasons (Ito, 2007).

What matters here is the issue of independence and autonomy of NGOs that implement assistance projects with funding from the government. In Western society, where the governments have been funding NGOs for decades, the issue has been discussed for a long time[4]. In fact, today a number of countries are actively involved in providing bilateral aid through NGOs for effective implementation of assistance projects for the recipients[5]. According to DAC, the percentage of resources channeled through NGOs in the total Official Development Assistance (ODA) is more than 10 percent in 10 out of 22 major countries[6]. That is, allocation of part of ODA for NGOs is becoming a trend in many donor countries (Nagasaka, 2007).

The situation surrounding NGOs in the Netherlands[7] is drawn upon as an example. Nagasaka (2007) reports that the government introduced, at an early stage, systems in which several domestic NGOs are allocated 10 to 20 percent of the total ODA budget and  implement ODA projects for the government. He also explains that NGOs autonomy is not affected by government policies. Under such circumstances, NGOs generally recognize that their independency and autonomy are secured despite the fact the government is their sponsor that provides them with a large amount of funds. In the meantime, the Dutch government acknowledges that NGOs possess networks in recipient countries through which assistance projects can be implemented effectively. Additionally, Dutch NGOs generally have good reputations in recipient countries.

The important thing is whether NGOs succeed in implementing activities in line with their mission goals and whether their goals and objectives are achieved, not by whom funds were granted for their activities and administration. This viewpoint is highly related to the issue of NGO accountability – to explain the significance,

effects, and progress of their assistance activities and projects to stakeholders, and to gain their understanding. As mentioned above, Japan has relatively a short history of government funding for NGOs, and it is only recently that Japanese NGOs started to increase their fiscal scale and influence in society. The issue of NGO accountability has not been sufficiently discussed in Japan.

Based on these points, the aims of this article are: 1) to review various concepts and relevant factors regarding accountability of civic organizations; and 2) to consider in a comprehensive way, by applying a case study, how Japanese NGOs can assume accountability. This article is structured as follows. Section 1 reviews preceding studies and focuses on the concept of accountability and the general functions and roles of NGOs in society. Section 2 introduces a case of evaluation of a project implemented by a Japanese NGO. Section 3 analyzes whether and how accountability was ensured in the case, and considers the analysis results. Finally in conclusion, future prospects and issues regarding NGO accountability in Japan will be given.

## 1. NGOs and Accountability

### 1.1 Concepts of Accountability

*Accountability* is "…a complex and abstract concept"(Brett, 1993). In Western society, there have been many studies regarding accountability of civic organizations including NGOs in international development since the middle of 1990s (Edwards & Hulme, 1996). For example, Edwards and Hulme (1996) state that "…accountability is generally interpreted as the means by which individuals and organizations report to a recognized authority (or authorities) and are held responsible for their actions."Kearns (1996) perceives the term in a broader concept by defining it as the means by which individuals and organizations meet stakeholders' various expectations regarding performance, responsiveness, and morality.

*Accountability* is often interpreted as "responsibility for explaining"in the Japanese context, and compared to *responsibility*. JANIC (2003) explains that both are concepts regarding judgment and recognition of the scope and level of fulfillment of a duty, and that the essential difference of the two lies in the fact that the former is relative, being founded on the relationship between two parties, whereas the latter is rather subjective and self-recognizing.

All these above-mentioned interpretations commonly focus on the relationship between the civic organization as a main entity and another party. This viewpoint will play a critical role when discussing NGO accountability.

### 1.2 Three Sectors and Accountability

A social system is generally composed of: 1) the government (public sector); 2) private enterprises (private sector); and 3) civic organizations (voluntary/non-profit sector)[8]. Several researchers have attempted to reconsider the position and roles of civic organizations by re-categorizing these sectors. Korten (2000) mentions that the above-mentioned three sectors are all essential in society, and that each entity has functions absolutely necessary to meet the needs of the group of people who do not belong to the party. The author also argues that the grass-root organizations (the fourth sector), possessing the functions or characteristics of all other three sectors, such as authority, economy, and collectivity, must be clearly distinguished from other types of civic organizations. The author concludes that today a number of organizations have become complex and are acting as intermediate organizations crossing the boundaries of the four sectors.

In the meantime, Nagasaka (2007) argues that, in order to develop democracy in Japanese society, the roles of civic organizations will be extremely important. The author insists that it is desirable to establish a social system in which the three parties－i.e. the government, private enterprises and civic organizations can build consensus and

Figure 1  Model of Consensus Building and Partnerships among Three Sectors
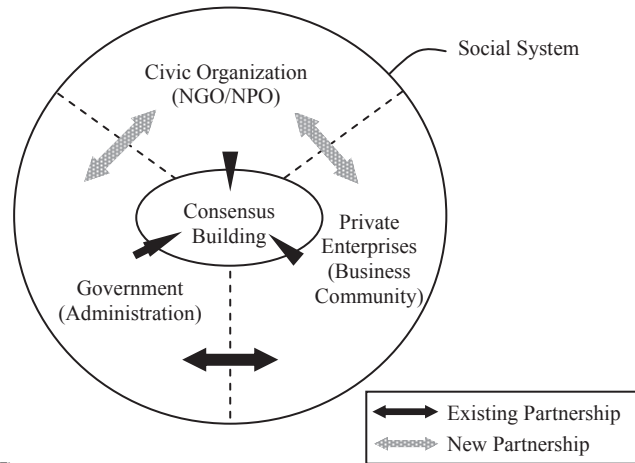


Source: The author, based on Nagasaka (2007)

execute fair administration based on equal partnerships. The author proposes a model of consensus building among the three sectors (Figure 1), and mentions that establishing partnerships between the government and civic organizations, as well as private enterprises and civic organizations is a current issue for the future. Also, regarding systems of ensuring accountability in each sector, the author points out that there exists no agency to objectively assess civic organizations regarding their accountability systems, whereas the government always must be accountable to citizens and is tested through fair elections, and private enterprises are evaluated by stockholders and consumers.
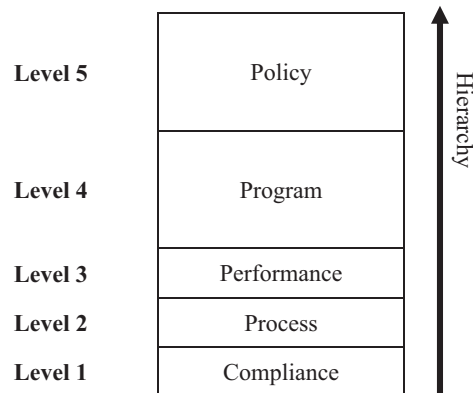
How do the government, private enterprises, and civic organizations differ in terms of the nature of accountability? Tanaka (2008) quotes a set of three models proposed by Brown (2003) focusing on stakeholders' relationships: 1) the Principal-Agent Model, commonly known as a corporate model, describing the relationship between the principal and the agent; 2) the Representative Model, describing the relationship between the voter and the government (politics and public administration) as a representative elected by the voter; and 3) the Mutual Model, describing equal relationships in which all stakeholders stand based on shared values. Tanaka (2008) points out that the third model is commonly preferred to the other two models by a number of civic organizations, which emphasize altruism and spontaneity.

Osumi (2002), on the other hand, quotes the hierarchy model of accountability, proposed by Stewart (1985), which responds to recent reform processes of the public administration systems (Figure 2). According to Osumi, in the traditional systems the accountability regarding 1) compliance and 2) process were more emphasized, however in recent new public management, the accountability on 3) performance, 4) program, and 5) policy are more valued, from the result-oriented perspectives focusing on outputs and especially outcomes. That is, there are different levels of accountability. This model can be applied when discussing accountability of private enterprises and civic organizations, however the accountability on policy is not questioned because these parties are not authorized to make public policy.

**1.3 Relationships between NGOs and Stakeholders in Development**
Here, the focus will be on NGOs tackling development issues. Korten (2000), with his experiences as the regional

Figure 2　Hierarchy Model of Accountability in Public Administration

| Level 5 | Policy |
| Level 4 | Program |
| Level 3 | Performance |
| Level 2 | Process |
| Level 1 | Compliance |

Hierarchy

Source: The author, based on Osumi (2002)

advisor on development management to USAID for Asia, divides the strategic patterns of NGOs into four generations based on the historical backgrounds, and considers the role of NGOs in each generation (Table 1). NGOs in the first generation, founded mainly from the seventeenth century towards the end of the Second World War, aimed to provide relief assistance to survivors of wars and natural disasters, however a number of NGOs later attempted to change their organization strategies emphasizing the concept of *development* because of the wide-spreading perception that such relief activities were only temporarily effective. NGOs in the second generation declared their strategies of assisting small-scale regional development through the introduction of resources, technology, etc. from outside the target community from the late 1970s. However, they were criticized because their strategy was based on

Table 1　Four Generations of NGO

| Generations | First | Second | Third | Fourth |
|---|---|---|---|---|
| Strategies & Aid Purposes | Rescue, Welfare | Community Development | Reforms to Sustainable Policy Environment and Social Systems | Inspiration for Social Movement by Citizens |
| Duration | Temporary | Project Period | 10-20 years | Indefinite |
| Target Range | Individuals, Families | Neighborhoods, Villages | Regions, Nations | Nations, Global Communities |
| Responsible Entity | NGOs | NGOs, Communities | All Relevant Public and Private Organizations | Networks of Citizens and Organizations |
| Main Financial Sources | Churches and Missions in the "North" | Public and Private Funds in the "North" | Public and Private Funds in the "South" | Public Funds in the "North" and "South" |
| Relationships with Stakeholders | Simple | | ⟶ | Complex |

Source: The author, based on Korten (2000)

the assumption of underdevelopment of the target community and the assistance that was provided promoted dependency of the beneficiaries. NGOs in the third generation, seen from the late 1980s, aim to create a fair and sustainable policy environment and social systems with strategies of realizing reforming policies and systems across individual communities. However, it has been pointed out that it requires a certain amount of time to bring necessary reforms to any system in the global society. With clear and more specific visions, NGOs in the fourth generation aim to realize larger-scale, people-centered development by declaring strategies of promoting voluntary social movements by the public at the global level. According to Korten (2000), the number of the fourth-generation NGOs is not large yet; however they have potential for the future in the sense that their action of promoting social system reform by the public can lead to resolving global-scale issues[9].

By looking at changes in strategic orientations across the four generations from a holistic viewpoint, it is clear that the way of providing assistance and involving the targets is changing－e.g. the duration of assistance becomes longer, the range of targets expands from individuals and families to communities, and from regions and countries to a global scale. Also, the entity of conducting assistance activities shifts from a single NGO to a mode of collaboration with communities and affiliated organizations, then to a mode of networks with the public. Furthermore, fund sources become diversified. It has, therefore, been pointed out that the relationships between stakeholders become more complex than before.

Hudock (2002), who served as a Democracy Fellow at USAID, categorizes NGOs into four groups from the viewpoint of their roles－that is, whether they are a proxy or an entity in charge of development. These four include: 1)「NGO in the north (developed countries)」founded mainly in Western society during and after the Second World War; 2)「NGO in the south (developing countries)」 founded as a recipient organization of financial support from an NGO in the north or founded on a voluntary basis; 3) 「NGO as mediator」 that assists another NGO or its target communities; and 4)「NGO as member organization,」 formulating a local social group that was founded for the purpose of responding to and resolving specific developmental issues. Hudock also refers to recent political and structural conflicts and tensions over the roles of NGOs between the south and north out of the four categories.

From the classification and categorization of NGOs proposed by the two authors, some essential questions are raised－whom NGOs represent for whose benefits in development; and how NGOs should ensure accountability while relationships between stakeholders become complex.

## 1.4 Reviews of Discussions on NGO Accountability

In respect to the questions above, Edwards and Hulme (1995) introduce the concept of NGO's *multiple accountability*. According to them, NGOs need to ensure accountabilities－ "upward"to donors, host governments, and their trustees as well as "downward"to their partners, beneficiaries, staff, and supporters. Shigeta (2007) and Tanaka (2008), however, point out that: 1) accountabilities to the two directions have not been well-balanced; and that 2) especially "downward"accountability has been somewhat ignored.

Edwards and Hulme (1995) also quote the difference between the two concepts of *functional accountability* and *strategic accountability*, which are proposed by Avina (1993). The former refers to NGOs' duty to report or explain regarding the organizations' resource use and immediate impacts in a short term, whereas the latter regarding the impacts that NGOs' actions have on the actions of other associations and the wider environment in a medium to long term.

Based on all these discussions, NGOs as a main entity are expected to equally ensure accountabilities to all stakeholders including donors, beneficiaries and local partners, and pay attention to the fact that the contents of reporting or explanation differ accordingly depending on the target stakeholder and the environment settings.

**1.5 Nature of NGO Accountability**

In conclusion, it becomes important for NGOs to first clarify how they interpret the significance of their activities, as well as their roles and raison d'etre in order to ensure accountability. The second key is how they view their relationships with various stakeholders including donors, supporters, beneficiaries, and partners. Finally, it is important to determine in what mechanism, about what contents and at what level they report or explain to each stakeholder respectively.

## 2. A Case of Evaluation of a Project Implemented by a Japanese NGO

In this and following chapters, a case of evaluation of a project implemented by NPO Kokkyo naki Kodomotachi (KnK)[10], to which the author of this article belongs, is drown upon and analyzed based on the argument in Section 1, for the purpose of considering the way Japanese NGOs ensure accountability.

**2.1 Outline of the Project and Evaluation Activities**

A massive earthquake claimed the lives of more than 5,000 people and injured 45,000 people on May 27, 2006 in Central Java, Indonesia, and KnK implemented its assistance activities for youths affected by the earthquake. Aiming to protect those youths and help them recover their mental stability, the association set up and run children centers in four villages in Bantul District, Yogyakarta Special Province where the level of damage was high, in partnership with the Indonesia Child Welfare Foundation (ICWF) as a local partner, funded mainly by the Japan Platform (JPF)[11]. At the centers, various activities were conducted, such as psychosocial care (recreational activities and psychological counseling), distribution of materials and food, and supplementary support for public school education. These activities targeted a total of approximately 400 children and youths ages five to fifteen. From the project formulation stage, the association had especially emphasized the importance of dialogue with the villagers, including village leaders and influential people, in order to design and implement activities considering the locality and community characteristics. Table 2 shows a summary of the project, including phases, durations and main donors.

Figure 3 indicates the existence of contracts[12] between KnK and major stakeholders with regard to the project and corresponding levels of accountability. For example, KnK signed a contract with JPF, and the contract scope ranges from achievement of the project goals to efficient and effective implementation of the project, and to compliance with and appropriate enforcement of regulations and accounting standards. Therefore, it corresponds to

Table 2  Outline of KnK's Project for Youths Affected by the Earthquake in Bantul District

|  | Phase | Project Duration | Main Donors |
|---|---|---|---|
|  | Initial Assessment | June 1 - 12, 2006 | Self-Financing (General Contributors) |
| 1 | Initial Response | June 13 - July 27, 2006 | JPF (Governmental Funds) |
| 2 | Emergency Assistance | July 28 - November 24, 2006 | JPF (Private Funds) |
| 3 | Recovery Assistance | November 29, 2006 - May 27, 2007 | JPF (Private Funds) |
| 4 | Recovery Assistance | May 28 – September 30, 2007 | Self-Financing (Private Enterprises and Associations, General Contributors) |

Source: The author

Figure 3  Contractual Relationships between Stakeholders and Levels of Accountability

| Written Contract | | Present | | | | Absent | |
|---|---|---|---|---|---|---|---|
| **Stakeholders** | | KnK | | | | | |
| | | ⬆ | ⬇ | ⬇ | ⬇ | ⬆ | ⬇ |
| **Accountability** | | **JPF** | **ICWF** | **CRC** | **Audit Firm** | **General Supporters** | **Target Community** |
| **Program** | Hierarchy | | | | | | |
| **Performance** | | | | | | | |
| **Process** | | | | | | | |
| **Compliance** | | | | | | | |

Note: The arrows between stakeholders indicate the flows of funds.
Source: The author

all four levels of accountability from "compliance" to "program." The association also signed the Memorandums of Understanding with: 1) ICWF and shared responsibilities regarding project implementation; 2) the Crisis and Recovery Center (CRC) of the Psychology Department of Gadjah Mada University regarding psycho-assessment (see later); and 3) commissioned an external Japanese certified accounting firm to make an audit of the project. Accordingly, the corresponding levels of accountability are indicated in relation with each stakeholder respectively. In the meantime, there existed no written contracts between KnK and target communities and general supporters. However, the project objective was commonly understood among these parties as the improvement of welfare for affected children and youths. Therefore, reporting and explanation on the implementation and effects of the assistance activities were requested (discussed later). Ensuring accountability at all the corresponding levels for each stakeholder is required.

In order to assess the effects of the project, especially Phases 1 to 3, three types of evaluations were conducted at the end of Phase 3.  Table 3 shows a brief summary of each evaluation.

"Project evaluation" was designed and implemented by JPF Secretariat as a donor, with an external expert involved, for the purpose of assessing the effects of each of the assistance projects for affected people in the areas hit by the earthquake, implemented by nine different Japanese NGOs respectively, all of which are members of NGO Units at JPF. Another important purpose of this evaluation was to ensure accountability to various stakeholders

Table 3  Outline of Three Types of Evaluation Activities

| Evaluation Activities | Field Survey Period | Planner (Category) | Evaluator (Category) | Evaluation Target | Evaluation Methods |
|---|---|---|---|---|---|
| **Project Evaluation/ Program Level Evaluation** | May 11 - 17, 2007 | JPF (donor) | JPF (donor) | Project (Phases 1-3)/ Program | Hearing from those concerned |
| **Psycho-Assessment** | February 25 - May 29, 2007 | KnK (Implementing Body) | CRC of the Psychology Dept. of Gadjah Mada Univ. (2 External Experts) | 40 youths of the beneficiaries | Observation, FGD, Hearing, Role-Play |
| **Impact Evaluation** | April 1 - May 27, 2007 | KnK (Implementing Body) | 8 ICWF staff members (local partner organization) | About 60 members of the community | Hearing from Community Members |

Source: The author, based on CRC (2007), ICWF & KnK (2007), JPF (2007)

regarding above-mentioned projects. All relevant projects[13] funded by JPF were placed in a program, and program-level evaluation was implemented. Also, each of these projects including KnK's was evaluated individually.

 "Psycho-assessment" was planned by KnK and implemented with the support of two CRC experts for the purpose of observing and assessing the changes that occurred to children and youths as beneficiaries through the project. As assessment targets, 10 youths who were showing traumatic symptoms[14] were arbitrarily selected in cooperation with the local staff in each of the four target villages respectively. For about three months, these children were assessed, in presence of KnK's Japanese staff, through behavior observation, hearing surveys, focus group discussions, and interviews with those concerned including their parents, guardians, the local staff, etc.

 "Impact evaluation" was planned by KnK and implemented in partnership with ICWF for the purpose of assessing changes in target communities brought by the project from the viewpoints of "living conditions" and "learning environment" of children in view of the locality and community characteristics. Prior to the evaluation, KnK's Japanese staff implemented a series of workshops on social survey methods and report-writing as part of capacity building for ICWF's local staff. In the actual evaluation surveys, the hearing method was mainly applied with the targets of 28 youths as project beneficiaries, 25 villagers including parents, guardians, school teachers, influential persons, etc. and 4 village leaders in the four villages.

## 2.2 A Summary of Evaluation Results
A brief summary of the results of the three types of evaluation is as follows.

### 2.2.1 Project Evaluation
Through Phases 1 to 3, the tent spaces at the centers were used effectively, and school furniture, equipment, supplies, and stationary were provided and distributed. Also, after-school supplementary classes were given for the purpose of improving the educational environment for affected children to be able to continue their school education. In addition, recreational activities focusing on self-expression, counseling, and extracurricular classes with a variety of themes were continuously offered at the centers. This contributed to the mental stability of the children and youths, as well as promoting their eagerness to study and their awareness of hygiene. On the other hand, better utilization of the knowledge of experts in psychosocial care was suggested as well as better coordination and consolidation with local governments and other relevant institutions[15] (JPF, 2007).

### 2.2.2 Psychological Evaluation
Right after the earthquake, a number of children and youths were being aggressive and were extremely sensitive to or afraid of noises. Some even lost their appetite or frequently cried. Half a year after they started to participate in various activities of the project, however, remarkable changes were seen in their behaviors, and the above-mentioned worrisome symptoms became less noticeable.

 From the collected data and analysis results, it turned out that participation in these various activities at the childrens centers positively affected the children and youths as beneficiaries. Although there were still a few children showing some worrisome symptoms at the end of Phase 3, most of the children were diagnosed to recover over time. Four children with serious symptoms, however, were considered to continuously need special treatment[16] (CRC 2007).

### 2.2.3 Impact Evaluation
Regarding the "living conditions" of children, most of them became mentally stable after starting to regularly participate in various activities at the centers. Also, they started to show their strong solidarity and harmony, and

obvious changes were observed in their awareness and behaviors in daily lives. Also, a number of children acquired manners through moral education classes at the centers, resulting in the improvement of attitudes at home, better relationships with adults around them, and more respect to the elderly. It also became evident that children who learned the importance of keeping themselves clean through hygiene education started to pay more attention to their living environment (ICWF & KnK 2007).

In terms of the "learning environment," it was observed that participation in after-school supplementary classes and sports activities at the centers helped quite a few children improve their academic and athletic performance, as well as feel more active and confident. These activities also significantly contributed to the general growth of many children in that their motivation in learning and reading remarkably improved (ICWF & KnK, 2007).

# 3. Analysis of Evaluation Case

In this section, the above-mentioned case of evaluation is analyzed in order to see whether accountability has been ensured regarding this project. The five-level accountability model proposed by Stewart (1984) and the concept of multiple accountability introduced by Edwards and Hulme (1995), as mentioned in Section 1, will be applied as a framework of analysis.

## 3.1 The Level, Target, and Contents of Accountability
### 3.1.1 Compliance and Process
Out of the five levels, the accountability on "compliance" and "process" focuses mainly on the compliance with regulations and accounting standards and appropriate enforcement of budgets, and efficiency and effectiveness of the project are not considered (Osumi, 2002).

Figure 4 indicates a conceptual model of each of the stakeholders to which KnK as the main entity should ensure accountability through Phases 1 to 3, as well as the directions of the accountability, each of which is explained as (a) to (n) as shown below. Regarding compliance and budget enforcement for the project, the organization submitted to (a) the donor a project completion report including the accounting report for each phase. In terms of the accounting reports, audits were implemented by an external certified public accountant, and the audit results have been submitted to and approved by (b') the donor[17]. These are all categorized as upward accountability. Therefore, there was a mechanism of ensuring accountability on compliance and process of the project outside the framework of the above-mentioned three types of evaluation.

### 3.1.2 Performance and Program
Regarding accountability on "performance" and "program," the main focus is on efficiency of project management at the output level, and relevance of goal-setting and the degree of goal attainment of the target program as a whole, or the degree of goal attainment at the outcome level in case of a single project in view of its placement in the entire program (Osumi, 2002).

As mentioned earlier, an impact evaluation was planned and implemented under KnK's supervision with ICWF's staff involved from the beginning. The implementation of evaluation surveys was directly linked to explanation and reporting to (c) ICWF on project effects and the degree of goal attainment. During the actual surveys, ICWF staff explained and reported the effects and significance of the project to (d) the target communities, including children as beneficiaries, their parents and guardians, those concerned in the target villages and other villagers. The results of the evaluation surveys were also presented to those concerned in each village at the end of Phase 3. In

Figure 4　Diagram of Stakeholders and Directions of Accountability
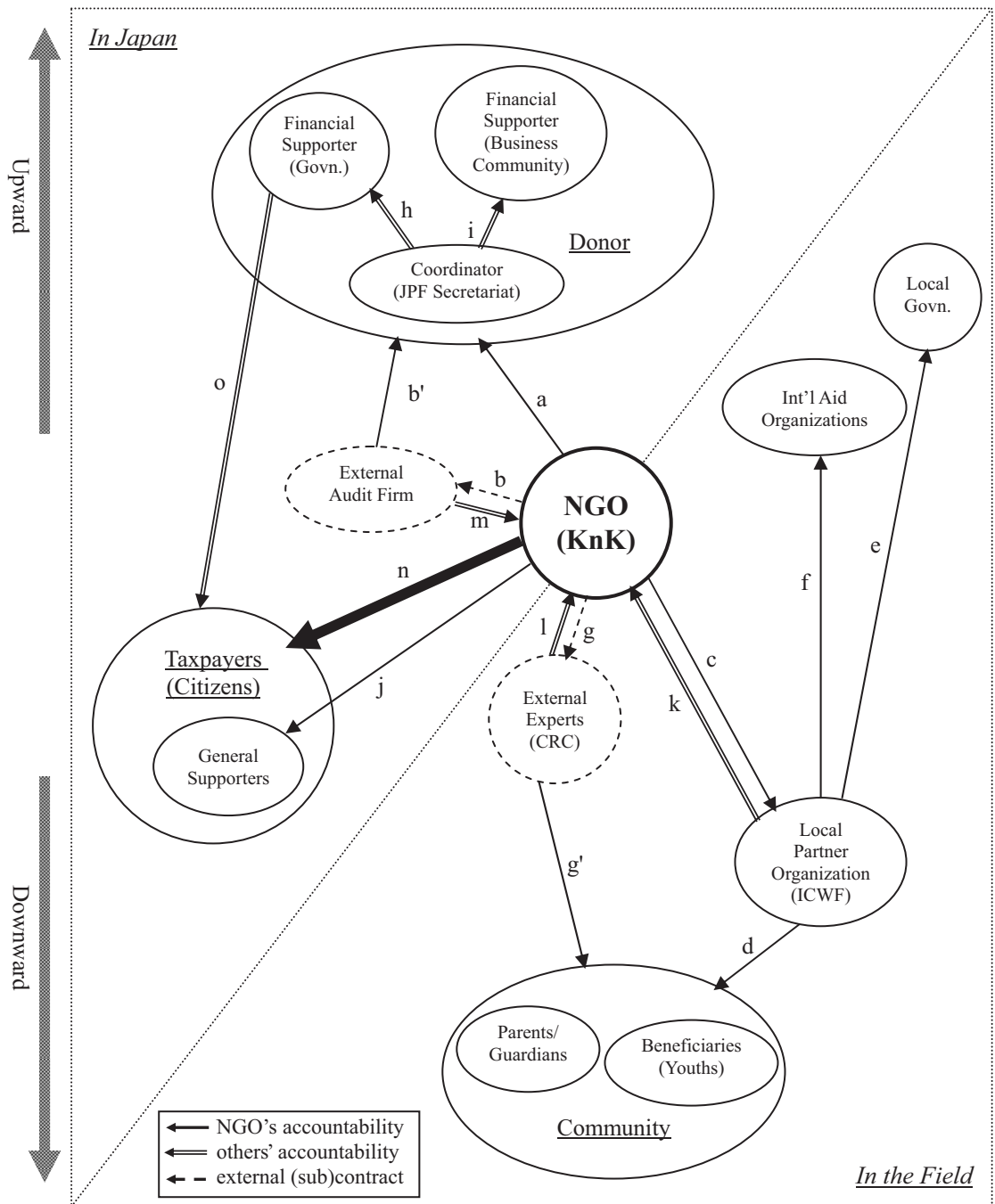


Note: Based on the concept of multiple accountability, the direction of accountability does not necessarily correspond to the flow of funds
　　(see Figure 3).
Source: The author

addition, these results were explained and reported to (e) the local government and (f) U.N. agencies and other aid organizations at cluster meetings of child and youth protection.

In the psycho-assessment, those children showing worrisome traumatic symptoms were mainly selected among the project beneficiaries and their psychological status were diagnosed by the experts under the supervision of KnK. The children's parents and guardians were involved in the survey. Upon completion, the results, influence and effects of the project on these children were explained and reported to (g') the community including the children, their parents and guardians by the experts in charge.

The results of the above-mentioned two types of evaluation were utilized to ensure downward accountability.

The project evaluation, in the meantime, was planned and implemented by JPF Secretariat in order to ensure accountability to different stakeholders as mentioned above. The results of evaluation of all relevant projects, as well as of the program were bound in a report and distributed. The results were also presented by JPF and some relevant NGOs to those concerned in (i) the business community as a donor at a meeting regarding these assistance projects. The author also attended the meeting and made a presentation about the results. This was a good occasion to utilize the evaluation results, and to report on or explain the significance and effectiveness of the project. Upon reporting the completion of the project, efficiency and effectiveness of the project, along with the results of the above-mentioned psycho-assessment and impact evaluation, were reported and explained to (a) JPF Secretariat, as well as those concerned from (h) the government and (i) the business community as donors to JPF. These are all categorized as upward accountability.

The progress and effects of the project were presented to (j) general supporters on KnK's homepage, at press conferences, in newsletters and annual reports, etc.

### 3.1.3 Vision and Strategy

As stated earlier, accountability on "policy" is not questioned to civic organizations, as they do not possess policy-making functions or authorities, unlike public administration. It is, however, important for a civic organization to decide how they should feed back their experiences, knowledge, lessons, and suggestions acquired through implementation of the project to the future visions and strategies of the organization and release the information to the public (discussed later). In this case, ensuring accountability to (n) the public including those concerned and supporters of the organization is required.

This project was the first children's center that KnK set up and ran in an area devastated by a large-scale natural disaster. The experiences and knowledge gained through implementation of the project were effectively used when projects to protect and support affected children were implemented in South Sumatra after an earthquake occurred in September 2007, and in the south of Bangladesh that was struck by a cyclone in November 2007. Other lessons learned from the project, such coordination with the local partner as an "NGO in the south," and project management tactics, including the way of establishing relationships with target communities, were utilized in shaping strategies for future projects of the organization. Information was also publicly released, including the organization's mission strategies that was established and published on its homepage in October 2007.

### 3.2 A Consideration on the Results of Case Analysis
### 3.2.1 Significance and Roles of Evaluation

The results of case analysis show that, in this project, a certain degree of accountability was ensured at all the five levels: compliance, process, performance, program, and strategy and vision (discussed later) and that the results of the three evaluations were effective in ensuring accountability. It was also suggested that it could be necessary to design different types of evaluations in a single project for different purposes and implement them in order to ensure

accountability to a number of stakeholders because the contents of explanation and reporting can vary depending on the stakeholder. As mentioned above, there exists no agency to objectively assess civic organizations regarding their accountability systems. Therefore, NGOs are expected to voluntarily take approaches to this issue, including making evaluation plans, establishing systems to ensure accountability without implementing evaluations, or utilizing information such as the results of monitoring and evaluations implemented by donors, and utilizing meetings and other opportunities provided by donors. It is also important to carefully note that the quality and objectivity of the results of evaluation implemented by different entities can affect the contents of accountability.

It also turned out that it is quite effective to involve the accountability target in the process of evaluation. In this project, ICWF staff came to understand in detail the project effects and the degree of goal attainment through the participation in the impact evaluation. In the psycho-assessment of children implemented in cooperation with CRC, their parents' and guardians' understanding of the significance and effects of the project was promoted by their active involvement in the evaluation. This consequently improved the level of transparency of the project and reliability of information. Direct involvement of stakeholders plays a significant role in ensuring accountability.

### 3.2.2 Development of NGO Accountability

Figure 5 indicates a comparison of differences in accountability expected of civic organizations and that of public administration, based on the hierarchy model of accountability proposed by Stewart (1984). The pyramid in the center shows that administrative functions and services provided by public administration are placed in the vertical formation of policy－programs－projects (Osumi, 2002). Out of the five levels of accountability expected of public administration, levels 1 to 3 correspond to the project level, level 4 to the program level, and level 5 to the project level. Since civic organizations do not possess policy-making functions or authorities, the accountability on "policy" is not questioned to civic organizations.

The author, however, proposes a concept of accountability on "vision and strategy" expected of civic organizations, instead of that on "policy" expected of public administration. As mentioned earlier, civic organizations are expected to feed back their experiences and knowledge to their visions and strategies and release the information to their supporters.

Figure 5  Comparison of Accountability between Civic Organizations and Public Administration



| | Accountability of Public Administration | Hierarchy | Levels of Administrative Functions and Services | | Accountability of Civic Organizations |
|---|---|---|---|---|---|
| Level 5 | Policy | | Policy | | **Vision and Strategy** |
| Level 4 | Program | | Program | | Program |
| Level 3 | Performance | | Project | | Performance |
| Level 2 | Process | | | | Process |
| Level 1 | Compliance | | | | Compliance |

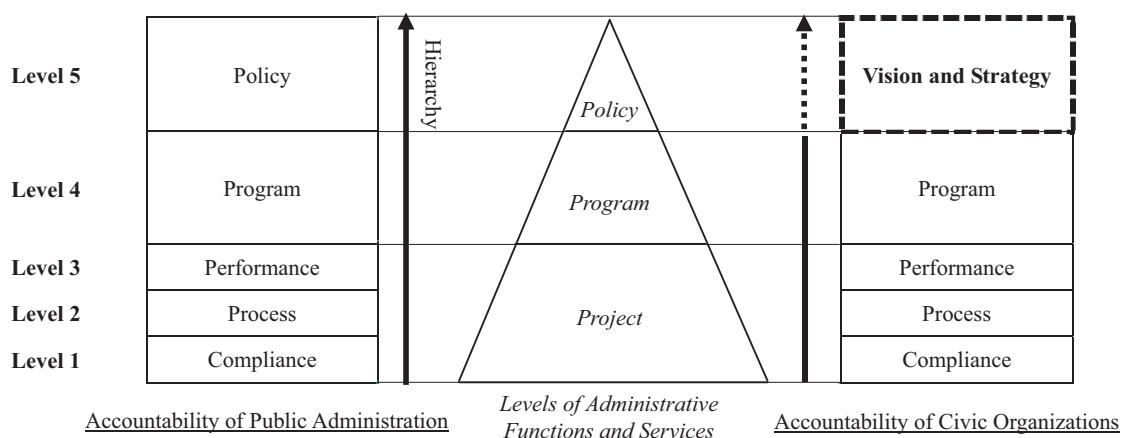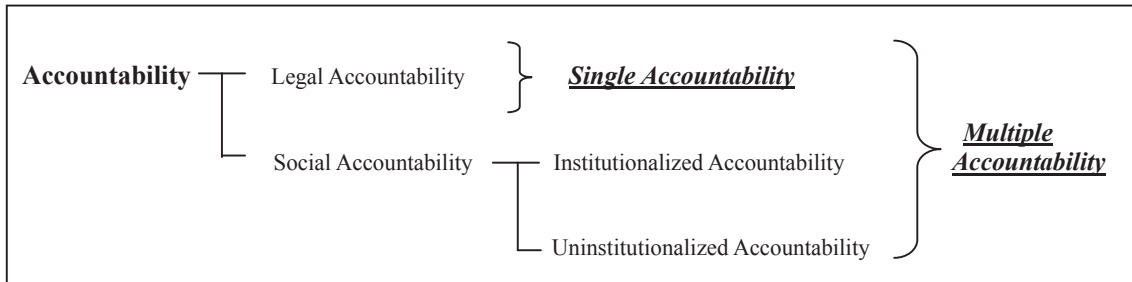Source: The author, based on Osumi (2002)

Figure 6  Classification of Accountability Concepts



Source: The author

Furthermore, from the viewpoint of the social responsibility assumed by civic organizations whose influence on society is increasing, it is required that civic organizations clearly present their missions and strategies to not only their supporters, but also the public, and embody their visions on what global societies should be in the future, as well as who they are through their missions and strategies. Accountability on visions and strategies will play a significant role in the future in the sense that it will be used to demonstrate the organization's capability in playing that role. This is a big issue for all civic organizations in Japan, including NGOs.

### 3.2.3 A Conceptual Classification of NGO Accountability

The evaluation case in this article is an example of applying the mutual model, one of three models proposed by Brown (2003) based on the concept of "multiple accountability." In this evaluation case, relationships with each of the stakeholders including the donor, the local partner organization and the beneficiaries, were focused under the common purpose of "improving welfare for children affected by the earthquake," regardless of the flow of funds, presence of written contracts, or upward or downward directions.

One point to keep in mind upon application of the model is to clarify whether each of the relationships with different stakeholders, for which ensuring accountability is assumed, is based on some written contract that can generate legal binding force, or on morality or social norms, as shown in Figure 3. These two are distinguished as "legal accountability" and "social accountability." Figure 6 shows the conceptual classification of accountability proposed by the author. "Accountability" conventionally refers to legal accountability, however multiple accountability also emphasizes social accountability. There are two categorizations in social accountability-one is institutionalized (e.g. (o) in Figure 4) while the other is uninstitutionalized (e.g. (d), (g') and (j) in Figure 4). It can be said that social accountability to downward stakeholders, especially uninstitutionalized accountability has often been ignored. NGOs are required to make efforts to ensure accountability to all stakeholders whether or not legal grounds exist in the relationships.

Figure 7 shows the author's re-definition regarding the directions in ensuring accountability based on the concept of multiple accountability proposed by Edwards and Hulme (1995). By assuming "neutral" (e.g. (d) and (g') in Figure 4) in addition to "upward" and "downward" directions from the viewpoint of capital flows, NGOs can have clearer perspectives in the attempt to become accountable for each stakeholder.

The contents of accountability expected of NGOs depend largely on the relationship between the NGO as the main entity and the stakeholder, although the relationship is always relative. It is, therefore, important for NGOs to be fully aware of their positions and roles and recognize whose benefits they represent upon implementation of a project so that they can establish an appropriate means to ensure accountability. For example, in this project case, KnK as an "NGO in the north" was the recipient of funds from the donor, and for ICWF as an "NGO in the south", KnK was

Figure 7  Re-definition of the Directions of Accountability
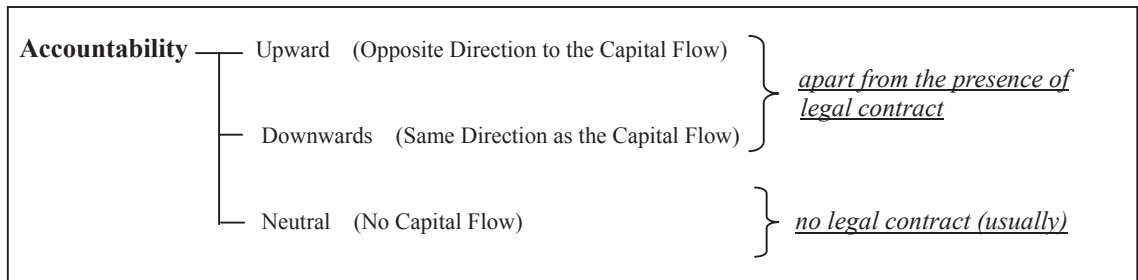
the direct fund provider. However, the project was implemented in equal partnership between ICWF and KnK, based on the recognition that experiences and knowledge of the local partner organization and experts were essential for psychosocial care for children and youths affected by the earthquake. This recognition eventually led to the designing and implementation of the impact evaluation in collaboration with ICWF and of psycho-assessment with the support of CRC, a local specialized institution.

In this project, (k) ICWF and (l) CRC (see Figure 4) are also expected to ensure accountability to KnK as a funds provider. ICWF was actively involved in the impact evaluation and reported the results and contents of the evaluation to KnK. ICWF also reported to KnK the results of the external audit for the project accounting implemented in Indonesia. CRC was also involved in the psycho-assessment and reported the contents of assessment results to KnK. These actions taken by the two institutions generally correspond the respective levels of accountability indicated in Figure 3, and it can be said that there was a mechanism to ensure accountability[18].


## Conclusion: Future Prospects and Issues regarding NGO Accountability in Japan

This article considered NGO accountability by applying the case of three types of evaluations conducted for a single project implemented by a Japanese NGO. The three following points will be the keys for NGOs to be accountable for their visions and strategies.

First is in terms of how to ensure NGOs' independence and autonomy. As mentioned in the beginning of the article, what is most important is whether NGOs succeeded in implementing their activities in line with their policies and goals, not by whom the funds were granted. The evaluation case introduced in this article was a good example of the mutual model, as cooperative relationships were observed between the NGO and each of the stakeholders. On the other hand, it is also true that recent high dependence on governmental funds by Japanese NGO can affect not only the organizations' independence and autonomy but also their *raison d'etre* and visions, as well as the establishment of their missions and strategic plans. In order to avoid such situations, it is necessary to explore diverse self-financial sources, independent of governmental funds.

Second is regarding how to realize visions and strategies. Based on the recent tendency of results-orientation in developmental assistance, NGOs are expected to present to the public, whenever necessary, how they will realize their visions and strategies by establishing their mission goals and strategic plans, clearly placing their projects in those frameworks, and appropriately grasping the degree of goal attainment and progress as well as relevant factors through performance measurement and evaluation.

Third is in regard to the public's involvement in evaluation. The evaluation case in this article demonstrates that

stakeholders' active involvement in evaluation processes was effective. Developing a mechanism to involve the public from outside in NGOs' project evaluation activities will promote further understanding of the public regarding target projects, as well as NGOs' visions and strategies. However, it is assumed that the projects are clearly placed in the frameworks of associations' mission goals and strategies.

In order to make a difference regarding these issues, NGOs need to take particular notice of the uniqueness in the environment surrounding civic organizations in Japan compared to Western society. First is regarding the nature of relationships between civic organizations and the government. Shigeta (2005) argues that the government generally has maintained strong power on society throughout the history, and that social services have been offered on the initiative of the government especially after the Second World War. He points out that, as a result, voluntary activities by the public have not been easily spread and a strong habitual reliance of the public on the government and public administration has been created in recent Japanese society. Nagasaka (2007) also argues that civic organizations have not easily developed as the concept of "kokyo" (public) was not clearly defined and society has been structured based on the dualism of "kou-shi" (official and private) since the national foundation in the Meiji era. It is still somewhat unpredictable as to how the relationship between civic organizations and the government will develop in the future.

Second is in terms of the trend of CSR[19] at private enterprises. According to Nakadera (2005), CSR has long been focused on due to its association with the trends of SRI[20], which became widespread in the 1980s in Europe and even earlier in the U.S. Various systems and guidelines have since been established by European governments, and in the U.S., private enterprises have taken the initiative in planning and implementing charitable activities in society. In Japan, on the other hand, private enterprises started to focus on the issue of CSR after 2000, and in spite of rapid development, a number of them are still seeking the way to assume CSR. For example, a number of cases of collaborative relationships of private enterprises with the governments or civic organizations are reported in Western society. Such cases are, however, still small in number in Japan, and especially civic organizations are not actively used.

Third is with regard to the public's consciousness and the degree of maturity of society. With the data published by the Center for Civil Society Studies (CCSS)[21] at the Johns Hopkins University in the U.S., Nagasaka (2007) points out that Japan's voluntary sector is still extremely immature as Japan's ratio of employees in civic organizations is the second lowest in the seventeen developed countries surveyed[22]. He also mentions that in Japan the citizen's participation in volunteer activities and willingness to do so are extremely low compared to in Western society by quoting data from the Cabinet Office (2005) indicating that less than 10 percent of citizens have actually participated in (Japanese) NPO activities and that 49 percent of them are unwilling to participate in NPO activities in the future. In addition, Ito (2007) points out that civic organizations have not yet been fully accepted in Japanese society. Much of these arguments are founded on the structure of Japanese society as mentioned above.

All these factors are related to the relationships between civic organizations and each of the composing members of society – i.e. the government, private enterprises, and citizens. These factors can hinder NGOs' ability to ensure accountability on visions and strategies. Therefore, NGOs need to be constantly aware of what each member expects from NGOs and what role these members expect NGOs to play. It is also important that NGOs raise awareness of people and establish new collaborative relationships with other members through active interaction with them. In addition, society as a whole needs to work on fostering and developing the subsystems of the voluntary sector in Japan, because the current subsystems are too fragile to support the voluntary sector as mentioned above, which cannot be easily resolved by civic organizations alone. It is expected that Japan's voluntary sector will further develop in the future, as active volunteer activities by citizens will help revitalize society and lead to creating new views of the world.

　　　The case applied in this article is categorized as emergency and development assistance rooted in the community. The areas of assistant activities and types of projects as well as relationships with stakeholders vary. Therefore, evaluation designs and implementation methods can be totally different depending on the objectives. Based on the implications derived in this article, it is strongly desired that further case analyses of NGOs' project evaluations will help accumulate practical knowledge and experiences on the issue of NGO accountability.

**Acknowledgement**

## Notes

1　The term "NGO" is generally used to refer to an organization of various types. Originally, the term was officially stipulated in Article 71 of the Charter of the United Nations signed in 1945, being used to refer to all non-profit organizations excluding the governments (Shigeta, 2005; Ito, 2007). A similar and related expression can also be found in the Charter of the International Labour Organization (ILO) established in 1919 (Charnovitz 2006). In Japan in general, "NGO" refers to a civic organization being engaged in international activities, whereas "NPO" refers to a civic organization domestically working in the areas of welfare and environment (Ito, 2007). Here, the term "civic organization" refers to an organization being active under a certain sustainable system, which is run mainly by citizens as members of society who are engaged with voluntary will and for no compensation. In this article, NGOs and NPOs are collectively referred to as civic organizations, and those engaged in activities in international cooperation are mainly discussed.

2　It refers to civic organizations that work on the improvement of welfare in developing countries as well as global-scale issues.

3　The percentages were calculated by the author based on the data provided by JANIC (2007). The Development Assistance Committee (DAC) also shows that the percentage of the government grant in the total NGO assistance of Japan in FY2003 reached as high as 35.9 percent (Shigeta, 2007).

4　For example, funds started to be granted to domestic NGOs by the International Cooperation Administration of the State Department, current Agency for International Development (USAID), in 1955 in the U.S.; by the Swedish International Development Cooperation Agency (SIDA) in 1962 in Sweden; by the Australian Agency for International Development (AusAID) in Australia and the Directorate of Development Cooperation, current Directorate-General for International Cooperation (DGIS), in the Netherlands in 1965; by the Canadian International Development Agency (CIDA) in 1968 in Canada; and by the Overseas Development Administration, current Department for International Development (DfID) in 1976 in the U.K. (Shigeta, 2007; Nagasaka, 2007).

5　In Canada, for example, in the 1980s, CIDA started granting funds to NGOs along with project proposals that were consistent with the priority of the policies. In recent years, the U.S. announced the policy to provide 40 percent of its Official Development Assistance (ODA) channeled through NGOs at the 1995 World Summit for Social Development, and has provided NGOs with project proposals (Nagasaka, 2007).

6　The percentage in Japan's ODA is 1.7 percent, which is the fourth lowest in 22 countries. The U.S. has not released its data (Nagasaka, 2007).

7　According to Nagasaka (2007), the country is often referred to and recommended among donor countries for its development cooperation and ODA policies, especially its collaborative relationships with NGOs. For example, DAC's peer review report on the Netherlands mentions that NGOs play a large role in the nation's ODA (OECD/DAC, 2006).

8　However, it is said that Japan is exceptional as the third sector refers to the consortium of the government and private

48 Satoshi Morita

# References

Avina, J. (1993). The evolutionary life cycle of non-governmental development organizations. *Public Administration and Development*, 13(5), 453-474.

Brett, E. (1993). Voluntary agencies as development organizations: Theorizing the problems of efficiency and accountability. *Development and Change*, 24, 87-118.

Brown, D. (2003). *A Framework for Strengthening NGO Accountability*. Manila: Asia Pacific Philanthropy Consortium.

Cabinet Office (2005). *NPO (minkan hi-eiri soshiki) ni kansuru yoron chousa (Survey on public opinions regarding NPO (nonprofit organizations)*. Tokyo: Cabinet Office.

Charnovitz, S. (2006). Accountability of Non-Governmental Organizations in Global Governance. In L. Jordan and P. Van Tuijl (Eds.), *NGO accountability－politics, principles and Innovations*. Sterling: Earthscan.

Crisis and Recovery Center (CRC), Faculty of Psychology, Gadjah Mada University (2007). *Psycho-assessment report－Evaluation of the effectiveness of the children center program in reducing children's trauma resulted from the 27th May 2006 earthquake of Yogyakarta*. Yogyakarta: Gadjah Mada University.

Edwards, M. and Hulme D. (1996). Introduction: NGO performance and accountability. In M. Edwards and D. Hulme (Eds.), *Beyond the magic bullet: NGO performance and accountability in the post-cold war world*. West Hartford: Kumarian Press.

Hudock, A. C. (2002). *Kaihatsu NGO to shiminshakai－Dairinin no minshuseiji ka? (NGOs and civil society: Democracy by proxy?)*. F. Nakamura and M. Tsuchiya (Trans.). Tokyo: Ningenno Kagaku sha (Original work published 1999).

Indonesian Child Welfare Foundation and Kokkyo naki Kodomotachi (ICWF & KnK) (2007). *Impact evaluation report: Assessment of the situation focusing on children's living condition and learning environment*. Tokyo: KnK.

Ito, M. (2000). Commentary: nihon no kokusaikyoryoku shimindantai (NGO) no genjo to tenbo (The present and future of civic organizations in international cooperation (NGOs) in Japan). In D. Korten, *Getting to the 21st century: Voluntary action and the global agenda*. T. Watanabe (Trans.). Tokyo: Gakuyo Shobo (Original work published 1990).

Ito, M. (2007). Dai 9 sho: Nihon shakai to NGO－Kokkyo o koeru shiminsoshiki (Chapter 9: Japanese society and NGOs: Civic organizations beyond borders). In N. Umahashi and A. Takayanagi (Eds.), *Gurobarumondai to NGO, shiminshakai (Global issues, and NGOs and civil society)*. Tokyo: Akashi Shoten.

Japan NGO Center for International Cooperation (2003). *NGO no akauntabiriti－JANIC akauntabiritii kijundukuri ni atatte (NGO Accountability: Creating JANIC's accountability standards)*. Tokyo: JANIC.

Japan NGO Center for International Cooperation (2007). *NGO deta bukku 2006－Suji de miru Nihon no NGO (NGO data book 2006: Statistics on Japanese NGOs)*. Tokyo: JANIC.

Japan NGO Center for International Cooperation (2008). *NGO dairekutori (NGO directory)*. Retrieved on April 28, 2008 from URL: <http://janic.org/modules/weblinks2/>.

Japan Platform (2007). *Jawato jishinhisaishashien hyokahokokusho (Evaluation report: Java earthquake program)*. Tokyo: JPF.

Japan Platform (2008). *Japan Platform towa (What is Japan Platform?)*. Retrieved on May 15, 2008 from URL: <http://w3.japanplatform.org/top.html>.

Kearns, K. P. (1996). *Managing for accountability: Preserving the public trust in public and nonprofit organizations*. San Francisco: Jossey-Bass Publishers.

Korten, D. (2000). *NGO to borantia no 21 seiki (Getting to the 21st century: Voluntary action and the global agenda)*. T. Watanabe (Trans.). Tokyo: Gakuyo Shobo (Original work published 1990).

Nagasaka, T. (2007). *NGO hatsu,＂shiminshakairyoku＂－Atarashii sekaimoderu e (＂Power of civil societies＂ generated by NGO: Creating new global models)*. Tokyo: Akashi Shoten.

Naikakufu (Cabinet Office of Japan) (2008). Tokuteihieirikatsudo sokushinho ni motoduku shinseijurisu oyobi ninshosuu, funinshosuu to (The numbers of applications, of certifications, and of certifications denied based on the Law to Promote

Nonprofit Activities). In *NPO homu peiji (NPO Homepage).* Retrieved on April 24, 2008 from URL: <http://www.npo-homepage.go.jp/data/pref.html>.

Nakadera, R. (2005). 1. Kigyo no shakaitekisekinin (CSR) wo meguru kokunaigai no doko (1. Overseas and domestic trends of corporate social responsibility (CSR)). In Global Environmental Forum (Ed.), *Kaihatsutojochiiki ni okeru kigyo no shakaitekisekinin: CSR in Asia—Heisei 16 nendo Wagakuni ODA oyobi minkankaigaijigyo ni okeru kankyoshakaihairyokyoka chosagyomu Part 1 (Corporate social responsibility in developing regions: CSR in Asia—FY2004 Research report on enhancement of environmental and social considerations in Japan*'s *ODA and overseas activities of private enterprises Part 1).* Tokyo: Global Environmental Forum.

Nippon Keidanren (Japan Business Federation) (2005). *CSR (Kigyo no shakaitekisekinin) ni kansuru anketochosakekka (Survey report on Corporate Social Responsibility (CSR)).* Tokyo: Nippon Keidanren.

Organization for Economic Cooperation and Development/Development Assistance Committee (OECD/DAC)  (2006). *Netherlands, DAC, peer review.* Paris: OECD.

Osumi, S. (2002). *Nyu paburikku manegimento—Rinen, bijon, senryaku (New public management: Philosophy, visions, and strategies).* Tokyo: Nippon-Hyoron-sha.

Shigeta, Y. (2005). *NGO no hatten no kiseki—Kokusaikyoryoku no hatten to sono senmonsei (History of development of NGOs: Development and expertise of international cooperation NGOs).* Tokyo: Akashi Shoten.

Shigeta, Y. (2007). Dai 7 sho: NGO to ODA (Chapter 7: NGO and ODA). In N. Umahashi and A. Takayanagi (Eds.), *Gurobarumondai to NGO, shiminshakai (Global issues, and NGOs and civil society).* Tokyo: Akashi Shoten.

Stewart, J. D. (1985). The role of information in public accountability. In A. Hopwood and C. Tompkins (Eds.), *Issues in public sector accounting.* Phillip Allan, Oxford: Humanities Pr.

Tanaka, Y. (2008). NGO no akauntabiritii to hyoka—Senryakuteki akauntabiritiiron kara (NGO accountability and evaluation: From the perspectives of strategic accountability). In K. Miyoshi (Ed.), *Hyokaron wo manabuhito no tameni (For those learning evaluation theories).* Kyoto: Sekaishisosha.

Wenar, C. and Kerig, P. (2000). *Development psychology: from infancy through adolescence.* 4th ed.  Singapore: Mc. Graw Hill.

【Article】

# Eight Models of Evaluation Use

## Go Yoshizawa

Graduate School of Public Policy, University of Tokyo

g-yoshizawa@pp.u-tokyo.ac.jp

## Abstract

A vast reservoir of knowledge and experience from utilization studies developed since the 1970s illuminates political and enlightenment functions in evaluation. This is virtually ignored by Japanese public agencies who report how they have used policy evaluation in a straightforward manner. As a bridgehead for applying the concept of evaluation use to evaluation studies and practices for public policy in Japan, this paper introduces independent/user-oriented, concrete/abstract (the process of use), substantive/symbolic (the purpose of use) and findings/process (the object of use) discriminators. With these discriminators and the appreciation of studies by policy analysts such as Weiss and Whiteman, this study illustrates eight different models of evaluation use with a novel, comprehensive matrix. Apart from the conventional typology of use including instrumental, conceptual and symbolic, this matrix gains a balanced perspective on the substantive and political nature of evaluation use.

## Keywords

evaluation utilization, utilization studies, substantive use, symbolic use, politicization of evaluation

## 1. Introduction

How does the research and practice of evaluation help society? In Japan, for instance, the Basic Guideline for Policy Evaluation (Cabinet Decision, December 16, 2009) says "it is necessary for individual administrative agencies to adequately utilize the result of policy evaluation as important information in the design and planning of policy and appropriately reflect it on the policy concerned". Based on this, each agency reports how they have used policy evaluation.[1] However, such reports are rather simplistic and only few of them are academically conducted independent from government authorities. By the same token, there have been very few reflexive analyses of evaluation studies and practices in Japan. Evaluation activities are strategically directed toward decision-making and problem-solving and they acquire specific knowledge relevant to the political context (Alkin and Taut 2003). Therefore, we must discern the appropriate use of evaluation more carefully than that of other social knowledge and academic studies. This is why the social needs for evaluation utilization studies have increased in recent years (Alkin 2003). A trendy concept called reflexive governance (Voß, Bauknecht and Kemp 2006) emerges from modern social

science studies, according to which we should pursue an approach in ways that evaluation studies contribute to the social betterment.

For this purpose, the present paper focuses on and discusses the use of evaluation. To classify use, the next section first introduces two dichotomies relating to the process of use - independent or user-oriented, and concretely used or abstractly used. It then introduces discriminators of whether evaluation is used with a substantive intention or with a symbolic intention (the purpose of use), and of whether the evaluation findings are used or the evaluation process is used (the object of use). The last section compiles these four dichotomies in a matrix and develops the discussion.

## 2. Taxonomy of Use

### 2.1 Independent/User-oriented

Firstly, let us elaborate whether evaluators have independent credibility or stand on the user side in the process from evaluation to decision-making. This may underlie the conflict structure of Weiss-Patton debate in the late 1980s (Alkin 1990). Roughly following both claims, Patton looked for a consensual and rational decision-making by providing users with useful information whereas Weiss insists that decision-making in a pluralistic and contentious community is irrational and political - what evaluators must do in such a context is to generate exact and adequate evaluation information (Smith and Chircop 1989; Shulha and Cousins 1997). In this way, Weiss requires evaluators to have independent credibility while Patton claims that evaluators must do evaluations standing on the user side. This distinction is not just epistemological, but as Patton (1997) proposes 'utilization-focused evaluation', is also made in methodological terms. It closely relates to the dichotomy introduced in the next section, where we develop the discussion further.

### 2.2 Concrete/Abstract

Use can be categorized according to whether the evaluation findings are used in a concrete way (e.g. decisions on the continuation of a program are made following the evaluation findings) or whether the evaluation findings are used in an abstract way (e.g. mid- and long-term policies are planned implicitly referring to the evaluation findings). These two - 'instrumental use' in the former and 'conceptual use' in the latter - are historically established concepts.

*Instrumental use* refers to directly citing or documenting information as a basis for action such as decision-making or problem solving that would not have been made otherwise (Rich 1977, p.200; 1978, p.101; Beyer and Trice 1982, p.598; Landry, Amara and Lamari 2001, p.336). Researchers who introduce this term depict the *engineering model* in which social research will provide empirical and conclusive evidence that helps to solve a policy problem (Gouldner 1956; Crawford and Biderman 1969, pp.235-237). The assumption in this model is that decision-makers have a clearer idea of their goals and a map of acceptable alternatives and evaluators answer specific requests for information and knowledge in a straightforward manner. Slightly different variations of this model have been termed the 'problem-solving' (Havelock 1969, 1971; Rein and White 1977), the 'decision-driven' (Weiss 1977; Nelson et al. 1987), the 'decisionistic' (Floden and Weiner 1978), the 'discrete functions' (Robertson and Gandy 1983), and the 'policy-driven' (Hanney et al. 2003) models.

Instrumental use does not always occur according to the engineering model. There is another model to understand the instrumental use in policymaking and administration - *technocratic model* (Robertson and Gandy 1983; Wittrock 1991). This model assigns independent credibility to evaluators, rather than being oriented to decision-makers as in the engineering model. The basic premise is that evaluation is designed and used to promote

technical solutions to political problems for decision-making (Fischer 1990, p.18), and evaluation can provide definitive and direct solutions. From this perspective, policymakers simply follow evaluators, professional recommendations (Foster 1980) and evaluation findings are automatically to be reflected on policy. The correspondences to this model in a wider domain of social science are referred to as the 'RD&D (research, development, and diffusion)' (Havelock 1969, 1971), the 'linear' (Cherns 1972), the 'knowledge-driven' (Weiss 1977), the 'empiricist' (Bulmer 1982), and the 'philosopher prince' (Coleman 1991) models.

However, the role of evaluation in politico-administrative contexts is many-sided, subtle and more complex than those implied by the term of 'instrumental use' (Weiss 1979; Wittrock 1991). The reality is that decisions often take shape gradually, without the formality of agenda, deliberation, and choice (Weiss 1980); this is *conceptual use*, which refers to influencing thinking about issues in order to gradually and indirectly shape decisions aimed at future policymaking (Rich 1977, p.200; 1978, p.101; Weiss 1980, p.382). Reported studies on the use of social research (Caplan, Morrison and Stambaugh 1975; Patton et al. 1977; Rich 1977; Lindblom and Cohen 1979; Weiss 1980; Whiteman 1985b, 1997) provide much empirical support for the proposition that conceptual use is more prevalent than instrumental use.

Conceptual use is often referred to in the *enlightenment model*, as contrasted with the engineering model (Crawford and Biderman 1969, pp.240-241; Janowitz 1970, pp.243-259; Bulmer 1982). Noteworthy is that this contrast has a double sense - whether the way of use is instrumental or conceptual, and whether evaluators have independent credibility or stand on the user side (see Table 2). The enlightenment model assumes that evaluation does not so much solve problems as provide an intellectual setting of concepts, propositions, orientations, and empirical generalizations that inform policy. No single evaluation has much effect, but, over time, concepts become accepted (Weiss 1978). This model is also called the 'limestone' model in the sense that information trickles like water through porous rock and exerts indirect and accumulative effects on policy change (Thomas 1985, pp.99-100).

The conceptual use for which evaluators are oriented to users is called the *classical bureaucratic model* (Wittrock 1991). In this model there is a need for legally and administratively trained personnel who are competent in the application of relevant rule systems but who have little or no need for any kind of social-scientific knowledge. Here the objective of evaluation is not to eliminate chance and uncertainty and uncontrollable social processes such as competition and other non-regulated forms of social interaction, but to deal with them and to live with them - according to the modern theory of bureaucracy and planning which is essentially just an extension of microeconomic reasoning.

### 2.3 Substantive/Symbolic

The distinction between instrumental use and conceptual use illustrated in the above was recognized, despite in a less clear form, by some researchers (Crawford and Biderman 1969) as early as the late 1960s. Since the concept of 'symbolic use' emerged from contexts of the politicization of evaluation in the late 1970s, three basic categories such as instrumental, conceptual and symbolic use (however these are termed) were established and have been referred to date in evaluation (utilization) studies (Pelz 1978; Young and Comtois 1979; Leviton and Hughes 1981; Beyer and Trice 1982; Shulha and Cousins 1997; Landry, Amara and Lamari 2001; Alkin and Taut 2003; Rossi, Lipsey and Freeman 2004, p.411). Symbolic use occurs when users use the findings of an evaluation exercise to legitimize and sustain their views and pre-determined positions (Weiss 1984; Pelz 1978; Knorr 1977). In terms of the above independent/user-oriented distinction, this is obviously the latter type of use. Symbolic use seems most likely with a reasonably clear user's commitment to an evaluation exercise, and relatively high (economic, scientific etc) uncertainty about a wide range of impacts of evaluation.

Table 1  Use of Evaluation Findings

|  | Substantive use | Symbolic use |
|---|---|---|
| Concrete use | Instrumental use | **Persuasive model** |
| Abstract use | Conceptual use | **Legitimizing model** |

Source: Author's own

Table 2  Models of Instrumental Use and Conceptual Use

|  | Independent | User-oriented |
|---|---|---|
| Instrumental use | **Technocratic model** | **Engineering model** |
| Conceptual use | **Enlightenment model** | **Classical bureaucratic model** |

Source: Adapted from Wittrock (1991, Table 1)

The reason why the above three categories have not been critically reviewed may be because evaluation studies and evaluation utilization studies have developed diverging from knowledge utilization studies and knowledge policy studies. For this the implication of symbolic use has been rarely reflected by researchers. An attempt at theory construction in which symbolic use is placed as a contrary concept of instrumental/conceptual use was developed chiefly by Whiteman, a policy scientist. This paper revaluates his achievement by referring to other relevant discussions.

Instrumental use and conceptual use illustrated in the previous section are based on the premise that there is an established division of labor between evaluation practitioners and users in scientific evaluation and the subsequent rational decision-making. This corresponds to the rational-objectivist approach proposed by van der Knaap and others (van der Knaap 1995; Albæk 1998; Dahler-Larsen 2000). However, evaluation is not a pure scientific endeavor but rather on a boundary between science and politics (Weiss 1973). Put it differently, users attribute scientific authority to their own decision and judgment but at the same time they maintain their political, administrative and legal control in the decision and judgment. This is the subjectivist-argumentative approach in evaluation and policymaking (van der Knaap 1995; Albæk 1998; Dahler-Larsen 2000), which describes the reality in which decision has to be made among conflicting values. Comparing to the premise of instrumental/conceptual use that the evaluation and decision-making process is objective and rational, symbolic use premises that the process is subjective and political. Accordingly, as Whiteman and others (Eaton 1962; Whiteman 1982, 1985a) pointed out, more persuasive would be to place symbolic use in another dimension and classify both instrumental use and conceptual use together as *substantive use*. Then the present study defines 'persuasive model' to correspond to instrumental use, and 'legitimizing model' to correspond to conceptual use.

As indicated in Table 1, substantive use includes instrumental use and conceptual use, and symbolic use includes persuasive model and legitimizing model.

*Persuasive model* occurs when evaluation becomes ammunition in political debate for whichever side finds its conclusions expedient and supportive. In concrete terms, it advocates issues, persuades people to act, neutralizes opponents, convinces waverers, and bolsters supporters (Weiss 1978, p.32; Johnson 1998, p.94). Since the chief function of such evaluation exercises is persuasion, users are perhaps more subject to the tendency to have the findings exploited for propagandistic aims (Merton and Lerner 1951). In this sense it is also labeled as 'political' (Boeckmann 1976; Weiss 1979). If users aim to facilitate persuasion, evaluation can be conducted in a way to support

Table 3  Models of Process Use

| | Substantive use | Symbolic use |
|---|---|---|
| Process use | **Interactive model** | **Tactical model** |

Source: Author's own

their favorite position rather than to provide credible evidence, and the use can serve the detriment of social betterment. A tendency not to persuade the public based on evidence but to compromise the evidence to facilitate persuasion is called the paradox of persuasive use (Henry 2000).

*Legitimizing model* is supporting or refuting predetermined positions and legitimizing decisions and resolutions based on political considerations such as political ideology, electoral hopes, coalition expediency or personal idiosyncrasies (Weiss 1977, p.15; Vedung 1997, p.275; Rossi, Lipsey and Freeman 2004, p.411). Evaluation findings are used selectively and often distortedly to publicly support a decision that has been taken on different grounds or that simply represents an opinion the decision-maker already holds (Knorr 1977). Unlike persuasive model, which helps users pressure other actors by providing evaluation findings directly, this model refers to a case in which users indirectly and implicitly use evaluation findings to legitimate and protect their own position (cf. Luukkonen-Gronow 1987). In other words, the difference is whether to explicitly persuade others to accept the users' position or to implicitly guide others to make it acceptable.

### 2.4 Findings/Process

All the aforementioned models of use can be grouped together as *findings use*, in that evaluation is understood as an outcome of the interests, needs, and calculations of particular actors (Patton 1997, pp.63-85). However, the process of the evaluation exercise can be used even if there are no outcomes of the exercise (Patton 1988). This is *process use*, which can be clearly distinguished from findings use because the objects of use are different. The concept emerged from evaluation utilization studies in the late 1980s (Shulha and Cousins 1997), and the taxonomy in which findings use and process use are contrasted becomes common (Smith 1988, pp.10-11; Kirkhart 2000, pp.9-10). Findings use occurs at summative (judgment-oriented) or formative (improvement-oriented) evaluation (Scriven 1972). On the other hand, process use is sometimes associated with evaluation in a constructivist fashion, which has been called 'fourth generation evaluation' (Guba and Lincoln 1989), in which evaluation and use are integrated (Finne, Levin and Nilssen 1995). Likewise, Dahler-Larsen (2000) illustrates that process use signals the integration of evaluation into routine regimes of organizational guidance and control as the third approach to evaluation, following the rational-objectivist and the subjectivist-argumentative. It should be noted here, however, that as we discuss below, process use has impacts on the development of personal competence or the substantive improvement of the quality of a project, so it does not always contribute to organizational management.

Patton, who introduced and has strenuously disseminated the term 'process use', appears to delimit process use as a substantive one. This may be related to a habit that he has carefully avoided using the term 'politics' in his discourse since the Weiss-Patton debate. By contrast, this study applies the term in a broader sense so as to include symbolic implications. Adopting Vedung's usage (1997, pp.274-276) and applying Whiteman's substantive-symbolic distinction (1982), as shown in Table 3, this study terms the substantive process use 'interactive model', as distinct from the symbolic process use hereafter referred to as 'tactical model'.

*Interactive model* brings about individual changes in thinking and behavior, and program or organizational changes in the evaluation process (Patton 1997, p.90). It occurs when users as individuals and groups involved in the

Table 4  Eight Models of Evaluation Use

|  |  | Substantive use | | Symbolic use | |
|  |  | Independent | User-oriented | Independent | User-oriented |
| Findings use | Concrete use | **Technocratic** | **Engineering** | X | **Persuasive** |
|  | Abstract use | **Enlightenment** | **Classical bureaucratic** | X | **Legitimizing** |
| Process use | Concrete use | **Interactive** | | X | **Tactical** |
|  | Abstract use | X | | X | X |

Source: Author's own

exercise learn by interpreting, understanding, and making sense of their experiences from the exercise process itself. The users apply evaluation-informed knowledge in conjunction with further research-based data and other forms of background like common sense, conventional wisdom, intuition, and the users' own first-hand experiences. Users' participation in the exercise can prompt both the clarification of standard operating procedures and their revision. This interactive model includes roles such as organizational learning, developing networks, extending communication, strengthening the project, and boosting morale (Forss, Rebien and Carlsson 2002; Patton 1997, 1998). Results of a survey on evaluation use suggest the perceived value of process use in that majority of respondents think highly of organizational development and empowerment or self-determination of individuals through internalizing evaluation processes (Preskill and Caracelli 1997).

There is a case where decision-makers use the evaluation process to signal to those concerned that something is being done about the problem. Important here is not the final outcome but the fact that the evaluation exercise is appointed and underway. Such use is called *tactical model*. For example, it refers to gaining time to postpone or neglect decisions and measures that should be taken (Knorr 1977, p.171; Vedung 1997, p.276). Decision-makers also use evaluation to deflect criticism, to try to avoid responsibility for unpopular policy outcomes, or to enhance the prestige of the agency involved (Weiss 1979, 1984; cf. Merton and Lerner 1951). In this mode, an evaluation exercise is also used to assume the pose of objective, scientific research (Suchman 1967, p.143). In any case, an empirical utilization study shows that tactical model has rarely occurred (Nilsson 1992).

**2.5 Summary and Discussion**

Individual classifications presented thus far are summarized in Table 4. Categories distinguishing between independent and user-oriented, concrete and abstract, substantive and symbolic, and findings and process are the eight models of use and indicated by boldface. As discussed above, symbolic use, by definition, does not give evaluators independent credibility. Thus all cells corresponding to "symbolic use - independent" are left unused.

As studies on the process use are relatively new, there has been little evidence on whether a distinction between instrumental (concrete) and conceptual (abstract) in the process use is possible, as a recent study (Alkin and Taut 2003) claims. There is concern about the conceptual process use is non-intentional so as to be called 'evaluation influence' rather than 'evaluation use'. It is also suspected that the tactical model is always concrete and never abstract. For these reasons, the present study tentatively assigns concrete use to interactive and tactical models (see Table 4). Future empirical studies will develop the discussion.

Features of this taxonomy distinct from existing studies are as follows: (1) it places symbolic use contrasting to instrumental/conceptual use in theoretical terms; (2) it defines persuasive/legitimizing model to contrast with instrumental/conceptual use; and (3) it introduces the substantive-symbolic distinction in the process use and

conceptualizes interactive model and tactical model. Furthermore, the present study illuminates individual two models each in instrumental use and conceptual use by introducing a distinction between independent credibility and user-oriented evaluation process and use.

Note that there are often cases where the evaluation use cannot be clearly classified into one model. In terms of the process of the use, the independent/user-oriented distinction should be much clearer, assuming that the procedures and institutions for the evaluation and use are different. In terms of instrumental use and conceptual use, these are not exclusive. Both can be applied as complementary at the same time (Marra 2000). The purpose of use can have substantive or symbolic intention, but the real use often includes both intentions (Feldman and March 1981; Pelz 1978). The use of evaluation findings and the use of evaluation process are clearly distinguishable as the objects of use are different. In this fashion, each model of evaluation use does not always take place independently from other models, but often times several models appear at once.

We have discussed some cases not clearly distinguishable between concrete and abstract, or substantive and symbolic use, but this does not mean that individual definitions and distinctions are ambiguous. Aside from remaining issues on the strictness in this conceptual taxonomy, inter-model or concurrent use can be observed in the reality. Nonetheless, there are possibilities that discourses about whether use is substantive or symbolic significantly vary by observer. For example, whilst a researcher judges an evaluation exercise as being politically used, decision-makers dismiss it or intentionally avoid from making such judgment. In brief, not only the use of evaluation can be political, but also discourses on the use can be political (argumentative). Future discussions on the politicization of evaluation should take into account such a reflexive perspective.

Dichotomies like independent/user-oriented, concrete use/abstract use, substantive use/symbolic use, and findings use/process use are not new in each but the taxonomy integrating these dichotomies is novel and original in the present study. As some literature (Weiss 1978, 1979; Vedung 1997; Owen and Rogers 1999, pp.110-113) only enumerates types of use without indicating a clear classification and distinction, this paper aims to arrange them. It is also necessary here to show a clear classification where some earlier taxonomic studies are confusing. For instance, Smith (1988, pp.10-11) proposes the directly-observable/perceptual distinction, which seems similar to the concrete/abstract distinction, but this taxonomy is mixing methodological and epistemological categories and less obvious and persuasive. Similarly, Johnson (1998, p.105) describes that behavioral use "mainly includes instrumental use, but may also include symbolic use, legitimative use, and action oriented process use" and "cognitive use includes cognitive oriented process use, enlightenment and conceptual use, and individual learning". Each of them includes various complex concepts and is very confusing.

Dichotomies not taken in the present study include immediate/long-term (Smith 1988; Kirkhart 2000), intended/unintended (Kirkhart 2000), partial or incremental/holistic (Smith 1988), and active/passive (Westerheijden 1997). The first two are omitted in order to distinguish 'use' from 'influence' and limit 'use' to less long-term and intentional consequence. In addition, there are relationships in which immediate and direct use loosely corresponds to concrete use and long-term and indirect use to abstract use, so it is better not to overlap classifications. The third distinction in the above, referring to the extent of use rather than the way of use, is outside of the scope of this paper. The last distinction is less appropriate in both pragmatics and typology because it is unlikely for external actors (and users themselves) to discern in a rational way whether user's consciousness is active or passive. Another reason is that the independent/user-oriented distinction performs a similar role to this - users can be more passive in the former and more active in the latter.

## 3. Conclusion

This paper picks up a relatively minor discipline called evaluation utilization studies and develops the discussion about what is the use of evaluation. Focusing on how evaluation is used, it analyzed the existing literature and made theoretical dichotomies from a comprehensive perspective concerning the process of use (independent/user-oriented, concrete/abstract), the purpose of use (substantive/symbolic) and the object of use (findings/process). These evaluation uses are then classified into eight models - technocratic, engineering, enlightenment, classical bureaucratic, persuasive, legitimizing, interactive and tactical. Practical and academic implications of this study are first that we come to understand the meaning of evaluation in more pluralistic and comprehensive terms. The deeper the understanding of conceptual use and process use scrutinized in the present study, the less myopic the practice of evaluation and decision-making. In addition, the academic articulation of symbolic use of evaluation in the government policy process makes evaluation in public policy more practical. Earlier evaluation studies tend to downplay political aspects by sticking to 'objective' or 'scientific' theories and methodologies of evaluation. Beyond the three categories - instrumental, conceptual and symbolic - as a still dominant typology in evaluation utilization studies, this paper acquires a balanced perspective on substantive and political implications in the use of evaluation, through Weiss's discussion on the politicization of evaluation and the revaluation of Whiteman's taxonomy. By accumulating findings of the use of evaluation in political and social contexts, the constructive relationship between evaluation and use will develop and evaluation studies toward better use of evaluation will flourish.

For a further study, it may be necessary to discuss where (for what), when evaluation is used in a decision-making and policymaking process and examine the relationship between the findings and the above models of use. The present case assumes that the use of public policy evaluation by bureaucrats, but it is required to theoretically and empirically illustrate to what extent this scope can be generalized. In addition to this, it is an important task to analyze why evaluation is used - factors affecting the use (i.e. inputs), contrasting to the present study on the use of evaluation (i.e. outputs). These factors include the nature of the issue, evaluator and user characteristics, process and product properties, evaluator-user interaction, information needs, legal and political context, and information processing style. From such input-output case analyses, we can develop studies on the characteristics of the use of evaluation in Japan and on the evaluation and decision-making process and institution.

## Note

1   For example, see Administrative Evaluation Bureau, Ministry of Internal Affairs and Communications, *Status of Policy Evaluation Implementation and Status of Its Policy Reflection* (in Japanese) (http://www.soumu.go.jp/s-news/2007/pdf/070608_1_1_H18-0301.pdf) and Cabinet Office, *Status Report on the Policy Reflection of Policy Evaluation Findings (FY2007)* (in Japanese) (http://www8.cao.go.jp/hyouka/h19hanei/houkokusho.pdf).

## References

Albæk, E. (1998). Knowledge, interests and the many meanings of evaluation: a development perspective. *Scandinavian Journal of Social Welfar*e, 7(2), 94-98.

Alkin, M. C. (1990). *Debates on Evaluation*. Newbury Park, CA: Sage.

―― (2003). Introduction. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International Handbook of Educational Evaluation*. Dordrecht: Kluwer Academic Publishers, 189-196.

Alkin, M. C. and Taut, S. M. (2003). Unbundling evaluation use. *Studies in Educational* Evaluation, 29(1), 1-12.

Beyer, J. M. and Trice, H. M. (1982). The utilization process: a conceptual framework and synthesis of empirical findings. *Administrative Science Quarterly*, 27(4), 591-622.

Boeckmann, M. E. (1976). Policy impacts of the New Jersey Income Maintenance Experiment. *Policy Sciences*, 7(1), 53-76.

Bulmer, M. (1982). *The Use of Social Research: Social Investigation in Public Policy-Making*. London: George Allen & Unwin.

Caplan, N., Morrison, A. and Stambaugh, R. J. (1975). *The Use of Social Science Knowledge in Policy Decisions at the National Level: A Report to Respondents*. Ann Arbor, MI: Center for Research on Utilization of Scientific Knowledge, Institute for Social Research, University of Michigan.

Cherns, A. B. (1972). Models for the use of research. *Human Relations*, 25(1), 25-33.

Coleman, D. A. (1991). Policy research - who needs it? *Governance: An International Journal of Policy and Administration*, 4(4), 420-455.

Crawford, E. T. and Biderman, A. D., eds. (1969). *Social Scientists and International Affairs: A Case for a Sociology of Social Science*. New York: John Wiley & Sons.

Dahler-Larsen, P. (2000). Surviving the routinization of evaluation: the administrative use of evaluations in Danish municipalities. *Administration & Society*, 32(1), 70-92.

Eaton, J. W. (1962). Symbolic and substantive evaluative research. *Administrative Science Quarterly*, 6(4), 421-442.

Feldman, M. S. and March, J. G. (1981). Information in organizations as signal and symbol. *Administrative Science Quarterly*, 26(2), 171-186.

Finne, H., Levin, M. and Nilssen, T. (1995). Trailing research: a model for useful program evaluation. *Evaluation*, 1(1), 11-31.

Fischer, F. (1990). *Technocracy and the Politics of Expertise*. Newbury Park, CA: Sage.

Floden, R. E. and Weiner, S. S. (1978). Rationality to ritual: the multiple roles of evaluation in governmental processes. *Policy Sciences*, 9(1), 9-18.

Forss, K., Rebien, C. C. and Carlsson, J. (2002). Process use of evaluations: types of use that precede lessons learned and feedback. *Evaluation*, 8(1), 29-45.

Foster, J. L. (1980). An advocate role model for policy analysis. *Policy Studies Journal*, 8(6), 958-964.

Gouldner, A. W. (1956). Explorations in applied social science. *Social Problems*, 3(3), 169-181.

Guba, E. G. and Lincoln, Y. S. (1989). *Fourth Generation Evaluation*. Newbury Park: Sage.

Hanney, S. R., Gonzalez-Block, M. A., Buxton, M. J. and Kogan, M. (2003). The utilisation of health research in policy-making: concepts, examples and methods of assessment. *Health Research Policy and Systems*, 1(2). Available at http://www.health-policy-systems.com/content/1/1/2.

Havelock, R. G. (1969). *Planning for Innovation through Dissemination and Utilization of Knowledge*. Ann Arbor, MI: Center for Research on Utilization of Scientific Knowledge, Institute for Social Research, The University of Michigan.

⸺(1971). The utilisation of educational research and development. *British Journal of Educational Technology*, 2(2), 84-98.

Henry, G. T. (2000). Why not use? In V. J. Caracelli & H. Preskill (Eds.), *The Expanding Scope of Evaluation Use*. San Francisco, CA: Jossey-Bass, 85-98.

Janowitz, M. (1970). *Political Conflict: Essays in Political Sociology*. Chicago: Quadrangle Books.

Johnson, R. B. (1998). Toward a theoretical model of evaluation utilization. *Evaluation and Program Planning*, 21(1), 93-110.

Kirkhart, K. E. (2000). Reconceptualizing evaluation use: an integrated theory of influence. In V. J. Caracelli & H. Preskill (Eds.), *The Expanding Scope of Evaluation Use*. San Francisco, CA: Jossey-Bass, 5-23.

Knorr, K. D. (1977). Policymakers' use of social science knowledge: symbolic or instrumental? In C. H. Weiss (Ed.), *Using Social Research in Public Policy Making*. Lexington, MA: Lexington Books, 165-182.

Landry, R., Amara, N. and Lamari, M. (2001). Utilization of social science research knowledge in Canada. *Research Policy*, 30(2), 333-349.

Leviton, L. C. and Hughes, E. F. X. (1981). Research on the utilization of evaluations: a review and synthesis. *Evaluation Review*, 5(4), 525-548.

Lindblom, C. E. and Cohen, D. K. (1979). *Usable Knowledge: Social Science and Social Problem Solving*. New Haven and London: Yale University Press.

Luukkonen-Gronow, T. (1987). Scientific research evaluation: a review of methods and various contexts of their application. *R&D Management*, 17(3), 207-221.

Marra, M. (2000). How much does evaluation matter?: some examples of the utilization of the evaluation of the World Bank's anti-corruption activities. *Evaluation*, 6(1), 22-36.

Merton, R. K. and Lerner, D. (1951). Social scientists and research policy. In D. Lerner & H. D. Lasswell (Eds.), *The Policy Sciences*. Palo Alto, CA: Stanford University Press, 282-307.

Nelson, C. E., Roberts, J., Maederer, C. M., Wertheimer, B. and Johnson, B. (1987). The utilization of social science information by policymakers. *American Behavioral Scientist*, 30(6), 569-577.

Nilsson, K. (1992). Utilization patterns and strategies in three policy sectors. *Science Studies*, 5(1), 29-46.

Owen, J. M. and Rogers, P. J. (1999). From evaluation findings to utilisation, *Program Evaluation: Forms and Approache*s. Sage, 105-131.

Patton, M. Q. (1988). Six honest serving men for evaluation. *Studies in Educational Evaluation*, 14(3), 301-330.

——(1997). *Utilization-Focused Evaluation: The New Century Text 3rd edition*. Thousand Oaks, CA: Sage.

——(1998). Discovering process use. *Evaluation*, 4(2), 225-233.

Patton, M. Q., Grimes, P. S., Guthrie, K. M., et al. (1977). In search of impact: an analysis of the utilization of federal health evaluation research. In C. H. Weiss (Ed.), *Using Social Research in Public Policy Making*. Lexington Books, 141-163.

Pelz, D. C. (1978). Some expanded perspectives on use of social science in public policy. In J. M. Yinger & S. J. Cutler (Eds.), *Major Social Issues: A Multidisciplinary View*. New York: Free Press, 346-357.

Preskill, H. and Caracelli, V. J. (1997). Current and developing conceptions of use: evaluation use TIG survey results. *Evaluation Practice*, 18(3), 209-225.

Rein, M. and White, S. H. (1977). Can policy research help policy? *Public Interes*t, 49, 119-136.

Rich, R. F. (1977). Uses of social science information by federal bureaucrats: knowledge for action versus knowledge for understanding. In C. H. Weiss (Ed.), *Using Social Research in Public Policy Making*. Lexington, MA: Lexington Books, 199-211.

——(1978). The use of science in policymaking: a comparative perspective on science policy. In D. E. Ashford (Ed.), *Comparing Public Policies: New Concepts and Methods*. Beverly Hills and London: Sage, 99-110.

Robertson, A. and Gandy, J. (1983). Policy, practice and research: an overview. In J. Gandy, A. Robertson & S. Sinclair (Eds.), *Improving Social Intervention: Changing Social Policy and Social Work Practice through Research*. London and Canberra: Croom Helm, 241-289.

Rossi, P. H., Lipsey, M. W. and Freeman, H. E. (2004). *Evaluation: A Systematic Approach 7th edition*. Thousand Oaks, CA: Sage.

Scriven, M. (1972). The methodology of evaluation. In C. H. Weiss (Ed.), *Evaluating Action Programs: Reading in Social Action and Education*. Boston, MA: Allyn and Bacon, 123-136.

Shulha, L. M. and Cousins, B. (1997). Evaluation use: theory, research, and practice since 1986. *Evaluation Practice*, 18(3), 195-208.

Smith, M. F. (1988). Evaluation utilization revisited. In J. A. McLaughlin, L. J. Weber, R. W. Covert & R. B. Ingle (Eds.), *Evaluation Utilization*. San Francisco, CA: Jossey-Bass, 7-19.

Smith, N. L. and Chircop, S. (1989). The Weiss-Patton debate: illumination of the fundamental concerns. *American Journal of Evaluation*, 10(1), 5-13.

Suchman, E. A. (1967). *Evaluative Research: Principles and Practice in Public Service & Social Action Programs*. New York: Russell Sage Foundation.

Thomas, P. (1985). *The Aims and Outcomes of Social Policy Research*. London: Croom Helm.

van der Knaap, P. (1995). Policy evaluation and learning: feedback, enlightenment or argumentation? *Evaluation*, 1(2), 189-216.

Vedung, E. (1997). *Public Policy and Program Evaluation*. New Brunswick and London: Transaction Publishers.

Voß, J.-P., Bauknecht, D. and Kemp, R., eds. (2006). *Reflexive Governance for Sustainable Development*. Cheltenham: Edward Elgar.

Weiss, C. H. (1973). Where politics and evaluation research meet. *Evaluation*, 1(3), 37-45.

⸺(1977). Introduction. In C. H. Weiss (Ed.), *Using Social Research in Public Policy Making*. Lexington Books, 1-22.

⸺(1978). Improving the linkage between social research and public policy. In L. E. Lynn, Jr. (Ed.), *Knowledge and Policy: The Uncertain Connection*. Washington, DC: National Academy of Sciences, 23-81.

⸺(1979). The many meanings of research utilization. *Public Administration Review*, 39(5), 426-431.

⸺(1980). Knowledge creep and decision accretion. *Knowledge: Creation, Diffusion, Utilization*, 1(3), 381-404.

⸺(1984). Increasing the likelihood of influencing decisions. In L. Ruttman (Ed.), *Evaluation Research Methods: A Basic Guide*. Beverly Hills, CA: Sage, 159-190.

Westerheijden, D. F. (1997). A solid base for decisions. *Higher Education*, 33(4), 397-413.

Whiteman, D. (1982). Congressional use of technology assessment. In D. M. O'Brien & D. A. Marchand (Eds.), *The Politics of Technology Assessment: Institutions, Processes, and Policy Disputes*. Lexington, MA: Lexington Books, 51-64.

⸺(1985a). Reaffirming the importance of strategic use. *Knowledge: Creation, Diffusion, Utilization*, 6(3), 203-224.

⸺(1985b). The fate of policy analysis in congressional decision making: three types of use in committees. *Western Political Quarterly*, 38(2), 294-311.

⸺(1997). Congress and policy analysis: a context for assessing the use of OTA projects. *Technological Forecasting and Social Change*, 54(2/3), 177-189.

Wittrock, B. (1991). Social knowledge and public policy: eight models of interaction. In P. Wagner, C. H. Weiss, B. Wittrock & H. Wollmann (Eds.), *Social Sciences and Modern States: National Experiences and Theoretical Crossroads*. Cambridge: Cambridge University Press, 333-353.

Young, C. J. and Comtois, J. (1979). Increasing congressional utilization of evaluation. In F. M. Zweig (Ed.), *Evaluation in Legislation*. Beverly *Hills & London*: Sage.

# Publication Policy of the Japanese Journal of Evaluation Studies

Last revised on 15th February 2005

The Purpose and the Name

1．The Japan Evaluation Society (hereinafter referred to as "evaluation society") publishes "The Japan Journal of Evaluation Studies (hereinafter referred to as "evaluation study") in order to widely release evaluation studies and outputs of practical activities to domestic and international academic societies, interested individual and institutions, and contribute to the advancement and prevalence of evaluation practice.

Editorial Board

2．The editorial board administrates editing of evaluation study based on the editorial policy stated below.

3．The editorial board is formed with less than 20 members of the evaluation society who are assigned by the board of directors. Terms of editors are two years but can be extended.

4．The editorial board assigns one editor-in-chief, two vice-editors-in-chief, and a certain number of standing editors among the members.

5．The editorial board may hold at least one meeting to discuss the editing policy, plans of editorial board, and others.

6．The editorial board reports activities to the board of directors as needed and receives approval. Also it is required to report the progress of the past year and an activity plan for the following year at the annual conference.

7．The editor-in-chief, the vice-editors-in-chief and the standing editors organize the standing committee and administrate editing on a regular basis.

Editorial Policy

8．The evaluation study, as a principle, is published twice a year.

9．The evaluation study is printed on B5 paper, and either in Japanese or English.

10．Papers published in the evaluation study are categorized as five types;

10.1．Review

10.2．Article

10.3．Research note

10.4．Report

10.5．Others

11．The qualified contributors are members of the evaluation society (hereinafter referred to as "members") and persons whose contribution is requested by the standing editors. Joint submission of members and joint submission of non-members with a member as the first author are accepted. Submission by the editors is accepted.

12．Submitted manuscripts are treated as the above categories, however, the standing editors will decide based on the application of the contributors and the following guidelines;

12.1．"Review" is a paper, which provides an overview of evaluation theory or practice. The editorial board will make the decision regarding publication.

12.2．"Article" is considered as a significant academic contribution to the theoretical development of evaluation or understanding of evaluation practice. The standing editors committee makes adoption judgments following the referee-reading process described in the next section.

12.3．"Research note" is a discussion equivalent to the intermediate outputs of a theoretical or empirical enquiry. The standing editors committee makes adoption judgments following the referee-reading process described in the next section.

12.4. "Report" is the study report related to a practical evaluation project or evaluation. The standing editors committee makes adoption judgments following the referee-reading process described in the next section.

12.5. "Others" includes requested papers for special editions organized by the editorial board and announcements from editorial board to members regarding publication.

13. The editorial board selects two referee readers. For the "article", the editorial board makes adoption judgments referring to the results from referee readings and comments provided by one editor assigned by the editorial board. For "review", "research note", "report" and "others", the editorial board makes adoption judgments referring to the results from referee readings.

14. When editors submit a manuscript, the editors are not allowed to attend any of the standing editors committee meetings or editorial board meetings regarding the manuscript.

15. The standing editors have alternative of approval or not-approval for adoption judgment of manuscripts submitted to any categories. However exception is permitted if the editorial board approves the publication after minor rewrite. Even if the manuscripts are considered insufficient as an "article", standing editors can decide whether the manuscripts are published as a "research note" or "report" if the authors wish to publish.

Formulation and Release of Submission Procedure

16. The editorial board formulates the submission procedure based on the editorial policy described above and release after approval from the board of directors.

Distribution

17. The evaluation study is distributed to all members for free and distributed to non-members for a charge.

Distribution of the Printed Manuscript

18. 30 copies of the respective paper are reprinted and distributed to the authors. The authors must cover any costs incurred by author's requests for printing more than 30 copies.

Release on the Internet

19. The papers published in the evaluation study are released on the internet with approval from the authors.

Copyright

20. Copyright of papers which appear in the evaluation study is attributed to the respective authors. Editorial right is attributed to the evaluation society.

Office

21. The office is in charge of administrative works for editing, distribution, and accounting.

# Information for Contributors
# (For English Papers)

Last revised on 29ᵗʰ September 2008

1．"The Japanese Journal of Evaluation Studies" is the publication for reviews, articles, research notes, and reports relating to evaluation.

2．"The Japanese Journal of Evaluation Studies" is primary published to provide opportunities for members of the Japan Evaluation Society (hereinafter referred to as "members") to exchange findings, and to contribute to further development of the study of evaluation both domestically and internationally. As a principle, this journal publishes the contributions submitted by the members. With the exception of requested papers, the first author must be a member. A submission (as the first author) is limited to one manuscript that has not been published or submitted in any form for another journal of academic association etcetera.

3．Adoption judgments of the manuscript are made at the discretion of the editorial board. Comments from two referee readers who are appointed for every manuscript are referred to in the screening process (the editorial board requests referee readers without notifying the author of manuscript).

4．Payment for the manuscript is not provided.

5．Papers published in "The Japanese Journal of Evaluation Studies" are released on the Internet at homepage of this academic society.

6．Regarding submission, manuscripts must be identified as one of the following categories: 1) article, 2) review, 3) research note, 4) report, and 5) others. However, the final decision of the category is made by the editorial board.
   "Article" is considered as a significant academic contribution to the theoretical development of evaluation or understanding of evaluation practice.
   "Review" is a paper which provides an overview of evaluation theory or practice.
   "Research note" is a discussion equivalent to the intermediate outputs of a theoretical or empirical study in the process of producing an "article".
   "Report" is the study report related to a practical evaluation project or evaluation.
   "Others" are manuscripts for special editions requested by the editing committee.

7．Manuscript Submission
   (1) Manuscripts may be written in either Japanese or English.
   (2) Correction by the author is only for the first correction.
   (3) English manuscripts should be submitted only after the English has been checked by a native speaker.
   (4) Submit four hard copies (A4 size) of the manuscript. Contact information including mailing address, telephone number, fax number, and e-mail address, and the category of the manuscript should be clearly stated.
       For approved manuscripts, after necessary rewriting, the author needs to submit two hard copies of the final paper as well as a text file saved on a DOS/V formatted floppy disk. Original figures, charts, and maps should be provided.
   (5) Total printed pages should not exceed 14 pages. Any cost incurred by printing more than 14 pages must be covered by the author.
   (6) The layout for English papers should be 30 mm of margin at left and right side, 10pt for font size, 43

lines on A4 paper (about 500 words per page). An abstract of 150 words should be attached to the front. 14 pages are equivalent to 7,000 words but the body should not exceed 6,000 words to allow for the title, header, figure, chart, footnotes, and references. Please note that the number of pages may be more than expected depending on the number of figures included.

8. Mailing address

Office of Japan Evaluation Society at International Development Center of Japan

Hitachi Soft Tower B 22nd Floor, 4-12-6, Higashi-Shinagawa, Shinagawa, Tokyo,

140-0002, Japan

Phone: +81-3-6718-5931, Facsimile: +81-3-6718-1651

E-mail: jes.info@idcj.or.jp

# Writing Manual of the Japanese Journal of Evaluation Studies (For English Papers)

Revised on 18th September 2002

1．Text, Charts, Figures, Graphs, Diagrams, Notes, and References

(1) The paper should be written in the follow order:

First page: Title; the author's name; Affiliation; E-mail address; Abstract (150 words); Keywords (5 words)

Second page: The main text; acknowledgement; notes; references
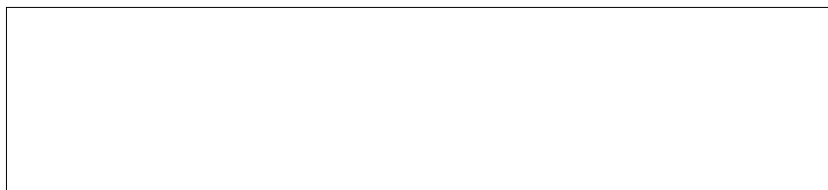
(2) Section of the text should be as follow:
1.
1.1
1.1.1
1.1.2

(3) Source of the charts, figures, graphs, and diagrams should be clarified. Submitted charts and others will be photoengraved, therefore it is important that the original chart is clear. Pictures shall be treated as figures.

Figure 1 Number of Students in the State of ○○

Note:
Source:

Table 1 Number of Accidents in the State of ○○

Note:
Source:

(4) Citation of literature in the text should be, (Abe 1995, p.36) or (Abe 1995).
(5) Note in the text should be, (------.[1] )

(6) Note and references should be written all together in the end.

Note

1 --------.

2 --------.

(7) Reference should list the literature in alphabet order, and arranged in chronological order. Follow the examples:

Book: author (year of publication). *Title of the book.* Published location: publishing house.

> (e.g.) Rossi, P. H. (1999). *Evaluation: A Systematic Approach 6ᵗʰ edition.* Beverly Hills, Calif: Sage Publication.

Article from magazine: author (year of publication). Title. *Title of the magazine*, volume (number), page-page.

> (e.g.) Rossi, P. H. (1999). Measuring social judgments. *American Journal of Evaluation*, 15(2), 35-37.

Article in Book: author (year of publication). Title. In editor (Eds.), *Title of the book.* Published location: publishing house, page-page.

> (e.g.) DeMaio, T. J., and Rothgeb, J. M. (1996). Cognitive interviewing techniques: In the lab and in the field. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research.* San Fransisco, Calif: Jossey-Bass, 177-196.

Book by two authors: surname, first name, and surname, first name. (year of publication). *Title of the book.* Published location: publishing house.

> (e.g.) Peters, T., and Waterman, R. (1982). *In Search of Excellence: Lessons from America's Best Run Companies.* New York: Harper & Row.

Book by more than three authors: surname, first name, surname, first name, and surname, first name. (year of publication). *Title of the book.* Published location: publishing house.

> (e.g.) Morley, E., Bryant, S. P., and Hatry, H. P. (2000). *Comparative Performance Measurement.* Washignton: Urban Institute.

(note 1) If some references are from the same author with the same publication year, differentiate by adding a,b,c as (1999a), (1999b).

(note 2) If the reference is more than a single line, each line from the second should be indented by three spaces.

> (e.g.) DeMaio, T. J., and Rothgeb, J. M. (1996). Cognitive interviewing techniques: In the lab and in the field. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research.* San Fransisco, Calif: Jossey-Bass, 177-196.

# Referee-Reading Guideline

1．Content of the Referee-Reading Guideline

This Referee-Reading Guideline is to provide explanation of the main publication judgment, procedure of the referee-reading, to the members who submit the manuscript and for the members who are requested to conduct referee-reading in order to carry out the procedure efficiently and effectively.

2．Purpose of Referee-Reading and the Responsibility of the Author

Referee-reading is necessary for the editorial board to make decisions of whether submitted manuscripts are appropriate to publish in the Japanese Journal of Evaluation Studies or not.

If there is doubt or obscurity identified in manuscripts during the referee-reading corrections may be required. Therefore, referee-reading also contributes to the improvement of the submitted manuscripts. However, although the manuscripts are requested corrections, the author is still solely responsible in regards to the contents and it is not attributed to the referee-readers.

Referee-readers are two persons who are requested by the editorial board depending on the specialty or the field of the submitted manuscript. People who are not members of this academic society also may be requested.

3．Items of Consideration in Referee-Reading

Five points are considered in referee-reading, however, the importance of each may be different depending on the type of manuscript.

(1) Importance and utility of the theme
(2) Originality of the study
(3) Structure of the logic
(4) Validity of verification and methodology
(5) Contribution to evaluation theory and practice

- For the article, all of above five are considered.
- For the research note, especially (1), (2), (3), and (4) are considered.
- For the report, especially (1), (3), and (5) are considered.
- For the review, especially (3) and (5) are considered.

4．Attentions in submission of manuscript

Besides above five viewpoints, basic completeness as a paper is also considered, for example;
- appearance of the paper is organized
- written according to the writing manual
- described simply and distinctive
- verification data is appropriately used
- notes and references are corresponding with the text
- terminology is appropriately used
- no wording and grammatical mistakes
- no errors and omission
- no punctuation mistakes
- expression in English abstract is appropriate
- word count is according to the manual

The above mentioned forms and contents are also considered. There have been cases in which graduate students and practitioners posted without organizing the manuscripts as a paper. On those occasions, referee-reading was not conducted. Necessary consultation is strongly recommended prior to submission.

5．Judgment Cases in Referee-Reading

(1) In the case of the manuscript which is considered acceptable for the publication but is not yet complete:
The referee reader should evaluate carefully whether the paper can contribute to the development of evaluation theory or evaluation studies.

- Verification is lacking but the theory and formulation are useful for academic development.
- Analysis lacking but useful for formation and promotion of new theory.
- The literature review is not of a high standard but, the overall study is meaningful.
- Comparative study is not up to standard but is meaningful as an example of application.
- Analysis is lacking but it is meaningful as an evaluation of socially and historically important cases.
- Analysis is lacking but it is meaningful as an evaluation of particular social activities.
- Organization and expression are not up to standard as a paper but the contents are worthy to evaluate.
- Logic is not strong enough but useful in practice.
- The paper has significance as a report.

(2) In case of the manuscript which is considered as difficult for publication:
- Awareness of the issue or setting of the problem is indecisive.
- Understanding or analytical framework of notion of basic terminology is indecisive or inappropriate.
- There is a lack in credibility of data for the grounds of an argument.
- There is no clear point of an argument or appropriateness of proof.
- Organization of the paper and presentation (terminology, citation, chart, etc) are inappropriate (or not consistent).

6．Judgment

The final decision will be made on publication at the standing editors committee following one of four patterns (listed below).  However, these judgments are not based on the number of errors but on the strength of the overall report.  In the case of (3) and (4), there is a possibility to be published as a different type of paper. If it is published as a different type of paper, major rewrite concerning the number of words may be required.

(1) The paper will be published.

(2) The paper will be published with minor rewrite.

(3) The paper will be published with major rewrite, however as a different type of paper (review, article, research note, or report).

(4) The paper will not be published; however there is the possibility that it will be published as a different type of paper (review, article, research note, or report).

# Correction Notice

The Japanese Journal of Evaluation Studies would like to announce a misprint in the last edition of the journal (Vol.8, No.2, March 2008). The piece entitled "An Evaluation Study of Program-Based Approaches (PBAs) for Development Assistance in the Primary Education Sector: A Case Study of the Education for All-Fast Track Initiative (EFA-FTI) in Honduras and Vietnam" by Satoko Miwa that was listed under the Research Note section is in fact classified as an Article and should have been listed under the Articles section. The JJES hereby apologizes to the author and our subscribers for any inconvenience or confusion caused.

*Japanese Journal of Evaluation Studies*