

Government Evaluations in Japan's Municipalities Based on Nationwide Survey Data

Yasutoshi Moteki

Graduate School of Humanities and Social Sciences, Hiroshima University

Nippon Urban Management and Local Government Research Association

E-mail: moteki@hiroshima-u.ac.jp

Abstract

Government evaluations in Japan are linked to business culture. This study investigated the government evaluations in cities and towns/villages using primary survey data. The surveys were conducted by the author and Hiroshima University in 2006, 2014, and 2015. The 2006 and 2014 surveys (Nationwide Survey of Japanese Cities Regarding Government Evaluations) were on city municipalities in Japan, and the 2015 survey investigated town and village municipalities (Nationwide Survey of Japanese Towns and Villages). This study, based on the survey data, intends to identify the changes in government evaluations in Japan's local municipalities over time and attempts to determine the differences among municipalities of different sizes.

A dominant method of government evaluations in Japan's municipalities is *Jimujigyo Hyoka*, a cross-sectional evaluation of programs via performance measures, which is implemented by 80% of municipalities with government evaluations. The survey data suggested the differences between the cities and smaller municipalities in terms of implementation and challenges. Cities were more likely than smaller municipalities to implement evaluations (84.6% vs. 37.4%). Some of the differences were with regard to implementation rates, the reasons for implementation, and the types of problems they were encountering. Thus, smaller municipalities needed different approaches of evaluation compared with cities.

An important finding of this study is that the smaller municipalities were relatively less interested in conducting government evaluations owing to their limited resources. Fiscal constraints tend to limit the number of public officials in small municipalities, and those officials tend to lack the expertise needed to conduct evaluations. Thus, officials in relatively small municipalities need focused training to obtain the skills necessary to implement evaluations.

Keywords: Government Evaluations, Local Municipalities, Urban Management, Public Officials, *Jimujigyo Hyoka*,

(1) Introduction

In Japan, *Jimujigyo Hyoka*, a cross-sectional evaluation of programs, is based on the method of performance measurement. This evaluation method focuses on the performance of each *Jimujigyo* (program and project) in all departments and creates evaluation results sheets in a unified format based on ex-post evaluations. Its use has spread throughout Japan's municipal governments under the influence of New Public Management (NPM), and, in particular, the reinventing government movement (Osborne & Gaebler, 1992) in the United States. Azuma (2002) proposed that fiscal deterioration, such as increases in long-term debt, was the background of NPM, arguing that emphasizing efficiency or effectiveness promoted the implementation of policy evaluation systems in Western countries.

There may also be a relationship between program evaluation and auditing. For example, Yamamoto and Watanabe (1989) explained performance auditing in Japan by linking it to deteriorating financial conditions. Hatry (2013) described the history of and relationship between program evaluation and performance measurement in the United States. Kudo (2015) employed literature reviews to analyze the association between conventional public administration and NPM theories, including post-NPM, New Public Governance (NPG), and administrative reforms. Yonezawa (2007) discussed evaluations in the context of reforms in higher education.

Theories and evaluation techniques became a focus for practitioners and scholars of Japanese public administration in the early 1990s. Morisugi (2000) examined manuals on evaluating transit projects such as roads, railways, airports, and seaports, and found that inconsistencies among the manuals across policy areas led to difficulties in preparing universal evaluation frameworks. Earlier, Tanaka (1989) had discussed how the Ministry of International Trade and Industry (MITI)

had evaluated national research and development projects in the 1960s through to the 1980s. It was not until the Mie prefectural government introduced *Jimujigyo Hyoka* and other evaluation measures that Japanese policy circles considered evaluation an important activity.

After the introduction of the system in Mie Prefecture, evaluations, referred to as *Gyosei Hyoka* (government evaluations), boomed in Japanese municipalities in the 1990s and improved governmental efficiency and effectiveness. *Gyosei Hyoka* also helped raise awareness of the importance of effective and cost-effective government policies among public officials. The concept of government evaluations as used in the United States means simply "evaluations by government." In contrast, *Gyosei Hyoka* in the Japanese context generally means "evaluations by local governments in relation to administrative reforms."

This definition of government evaluations in local government practice and research as well as the definition used by the Ministry of Internal Affairs and Communications (MIC) differ. First, the term *Gyosei Hyoka* is used in the title of the MIC's Administrative Evaluation Bureau (*Gyosei Hyoka Kyoku*). This *Gyosei Hyoka* of the MIC is often referred to as *Gyosei Hyoka-Kanshi* in Japanese, meaning administrative evaluations and oversight. Second, the MIC conducts a survey of the implementation of *Gyosei Hyoka* by local municipalities about every three years. The broad definition of *Gyosei Hyoka* used in the survey is taken from https://www.soumu.go.jp/main_content/000501750.pdf and reads as follows:

The term *Gyosei Hyoka* in this survey is defined as judging the relevance, achievement and results of policies, measures and administrative programs, regardless of whether they are implemented before, during or after the event, based on certain criteria and indicators. The scope

of policies to be covered corresponds to the scope of policies of 'policy evaluations' in the national government level and 'Administrative Program Review (*Gyosei Jigyo* Review)' of the central government.

These two uses differ from the concept of *Gyosei Hyoka* in the study of local administration and its practice. *Gyosei Hyoka* were often adaptations of the evaluation methods used by private enterprises. They tended to ignore the organizational differences between public and private entities. Japanese academic supporters of *Gyosei Hyoka* often introduced methods that had been successfully implemented by private enterprises into municipal governments without carefully considering the differences. The lack of attention to these differences may have had unintended harmful consequences, such as evaluation fatigue related to the large amount of paperwork involved in these activities.

This study analyzed the opinions of public officials managing municipal *Gyosei Hyoka* by using primary national-level survey data on implementation and problems related to government evaluations in cities and towns/villages. The national survey on towns and villages was conducted in 2015, and its data were compared with those of a survey conducted on cities in 2014 (Moteki, 2015). This paper discusses government evaluations in Japan at the municipal level using survey data gathered by the author in 2014 and 2015 in addition to the definitions and status of government evaluations in the Japanese context. This paper has four parts. A review of the literature on the evaluation methods used by Japan's municipal governments is followed by a description of the survey methods used in this study. A comparative analysis of the 2014 and 2015 surveys is presented next. The conclusion summarizes the differences between the cities and smaller municipalities, and I argue that smaller municipalities

have to tailor their evaluation methods to meet their needs and human resources.

(2) Literature review

Several quantitative studies of government evaluations in Japan have been conducted. Japan's Ministry of Internal Affairs and Communications conducts a mail survey (Status of Government Evaluations in Local Municipalities) of all municipal governments, including prefectures. These surveys are carried out once every three years, and the response rates are very high (e.g., the response rate in the 2016 survey was 100%). However, the reports are limited to tabulations of responses, although the raw data from the individual answer sheets of each respondent are available in Excel format (Ministry of Internal Affairs and Communications, 2017). The ministry has not analyzed trends, patterns, or other changes over time.

Tabuchi (2010) analyzed the 2009 survey data collected by Mitsubishi Research Institute. These surveys were carried out from 1998 until 2009 to clarify the state of municipal governments' administrative evaluations. Tabuchi (2010) summarized the survey findings across time and organized the evaluations into four stages, each lasting roughly three years. As the timing differed across municipalities, "the municipalities introducing the evaluation system are mixed on the four stages" (Tabuchi, 2010, p. 33). They faced various challenges, including finding ways to eliminate the burdensome feeling of engaging in the evaluation work, ways to use the evaluation results, and ways to move away from evaluations performed by the government alone.

Other papers and reports based on survey data mostly reported simple tabulations of the questionnaire item responses. The relationships between variables have not been statistically tested. Sato's (2013a) cross-sectional survey of Japan's municipalities in 2012 (excluding prefectures, towns, and villages) aimed to

clarify the structure and function of government evaluations for the administrative management system. The 810 municipalities included the Tokyo Metropolitan District and had a response rate of 73.8%.

According to Behn (2003), public managers measure performances in order to evaluate, control, budget, motivate, promote, celebrate, learn, and improve. Among these reasons, *Jimujigyo Hyoka* in municipal governments has been focused toward making improvements in the budget mostly because of financial crises. Other program evaluation methods (particularly the logic model) that are considered important by the American Evaluation Association (AEA) are not widespread in Japan.

Noutomi and Nakanishi (2007) analyzed the characteristics of NPM reform movement started in the mid-1990s in Japan's municipal governments by focusing on performance budgeting, total quality management, and target-based budgeting. They concluded that the performance measure used by the municipal governments was similar to conventional budgetary control. However, implementing a performance measure led to the diffusion of the notion of program evaluation, including logic models, which is an ongoing process. The Ministry of Internal Affairs and Communications recently added this to the national training in policy evaluations for government officials. Uchida (2009). Uchida (2009) discusses the disruptions that NPM-type government evaluations cause to small municipalities. This study's questionnaire of the survey also focuses on small municipalities and aims to identify differences in the situation between small and large municipalities.

(3) Government evaluations in Japan

Gyosei Hyoka is Japan's main evaluation tool. Although it is sometimes interchangeably used with *Seisaku Hyoka* (policy evaluation), the latter mostly indicates evaluation activities in the central government,

usually comprising scientific policy evaluations. *Seisaku Hyoka* has three aspects: the *Jigyo Hyoka* (project evaluation), the *Jisseki Hyoka* (performance evaluation), and the *Sogo Hyoka* (comprehensive evaluation) (Koike et al., 2007).

Gyosei Hyoka usually concerns municipal governments and is often about government downsizing or inefficiencies. It has two approaches: (1) *Jimujigyo Hyoka*, which uses benchmarking and policy indicators for public works' (infrastructure) evaluations; and (2) *Gyomu Tanaoroshi* (work process analysis), started by the Shizuoka Prefecture under the guidance of Professor Kitaoji of Meiji University in FY 2003. The dominant methods of evaluation used by the *Gyosei Hyoka* of municipal governments is the *Jimujigyo Hyoka* Cross-sectional Program Evaluation System.

Jimujigyo Hyoka became popular in Japan mainly in response to the United States' 1993 Government Performance and Results Act (GPRA) and the reinventing government movement (Osborne & Gaebler, 1992). The GPRA triggered a boom in government evaluation activities in the United States. Since the GPRA Modernization Act of 2010 (GPRAMA), federal departments and agencies have developed strategic plans, goals, and indicators to implement the GPRAMA. A system for evaluating and improving policies based on these indicators was introduced. In 1996, Mie Prefecture was the first Japanese municipality to implement *Gyosei Hyoka* based on the US experience, and other Japanese municipalities followed suit. Eventually, *Seisaku Hyoka* was implemented at the national level through the Government Policy Evaluations Act (Act No. 86 of 2001; Act 86).

Along with the law, policy evaluations by the central government have been governed by the Standard Guidelines for Policy Evaluation published by the Government Council for Policy Evaluation in 2001. The roles of each ministry's managing department are

described in Section 1 under Chapter 3. Two of the five roles are “planning and formulation of basic matters related to policy evaluation of administrative work in charge (development of implementation guidelines and management policies for evaluation implementation)” and “promotion of training and securing human resources capable of policy evaluation.” Training people to be policy evaluators is an important foundation for achieving effective evaluations; however, only one study on evaluators’ competencies has been conducted in Japan (Sato, 2013b). The Ministry of Internal Affairs and Communications regularly conducts trainings on policy evaluation methods for both national and municipal public officials. National-level trainings take place in Tokyo, and municipal-level trainings are held in regional central cities. It is not easy for municipal officials to participate because the venues, Tokyo or the central regional cities of each area, are located far from local municipalities.

Yamaya (2002) pointed to academic research and government practices while focusing on the influences of the evaluation methods used by private enterprises on government evaluations and argued that policy evaluation in Japan “has been ‘evaluation’ far from the model drawn by the theory of evaluation” caused by the lack of “interpersonal exchanges” between academics and practitioners (p. 337). Yokoyama (2009) proposed that when the policy evaluation system began in municipalities in the 1990s, its main purpose was administrative reform. She used *Nihon Keizai Shimbun* (a Japanese newspaper) data to analyze the relationships between implementation by municipalities and possible predictors such as the extent of administrative reform. Two of these variables were “the introduction of evaluation systems” and “making the balance sheet, evaluation results, and committee meeting minutes publicly available.” She found a statistical relationship between the independent variables and “improvement of administrative services.”

Gyosei Hyoka became closely linked to management reform and was significantly influenced by the management practices of private enterprises. Former automaker employees employed at government offices were often asked to help guide administrative reforms and improvements. Private enterprise management tools were directly applied to improve governmental operations, perhaps related to the NPM movement. Applying the practices of the private sector contributed toward the dissemination of *Gyosei Hyoka*, the gaps between the evaluations being used and program evaluation theory, and the evaluation fatigue and formalization (ritualization) of the systems.

Gyosei Hyoka by municipal governments often related to efficiency through *Gyosei Keiei* (public management) and *Gyosei Kaikaku* (administrative reform). However, it is not clear whether the evaluations benefitted the public officials or the stakeholders. Observations in the current study revealed that implementation was a heavy burden on public officials in the central evaluation departments because evaluations at the municipal level tended to be challenging for the officials on account of negative public opinion of government and public demand for efficiency. The public and stakeholders tended to mistakenly perceive evaluation as a means of achieving administrative reform by reducing waste in public expenditure. Some business management scholars persistently argue that business theory may easily apply to public administration, which encourages this type of misunderstanding. Ueyama (1998), considered as one of the most influential authors by practitioners and citizens, wrote “*Gyosei Hyoka*” no *Jidai* (*Era of Gyosei Hyoka*). The subtitle of this text is “Perspectives from Management and Customers.” In short, *Gyosei Hyoka* in Japan seems overly focused on efficiency and on downsizing administrative organizations. As Yamaya (2002) explained, *Gyosei Hyoka* is very different from the original notion of program evaluation accepted

around the world and the concept of “government evaluation” used by western academic societies, including the American Evaluation Association, the Canadian Evaluation Society, and the European Evaluation Society.

Many Japanese public officials have recently come to emphasize on *Hyoka Zukare*, a negative aspect of government evaluation, which means “evaluation fatigue.” This and another similar term, “evaluation exhaustion,” has come to be used by the Organisation for Economic Cooperation and Development (OECD, 1997). Blackmore (2003) argued that the problem of evaluation exhaustion exists among students as well. A previous study on program evaluations pertaining to international comparison focused on Sweden (Johnsen, 1999). Furthermore, there have been studies by Lahey and Nielsen (2013), who have presented the Canadian case, and Arthur et al. (2012), who have examined performance auditing in Norway.

Although many Japanese municipalities have evaluated programs based on *Jimujigyo Hyoka* under the name of *Gyosei Hyoka*, some practitioners and researchers have been pointing out at the negative effects concerning the costs of the systems and the side-effects on staff, such as evaluation fatigue (Sawada, 2010). Consequently, the logic model is attracting attention as a framework for visually grasping the logical relationships between the overall goals (mission) and the activities implemented (programs). Logic models are increasingly being used by municipalities to improve *Jimujigyo Hyoka*. For example, in FY 2013, the Ministry of Internal Affairs and Communications held lectures on logic models as a part of national training in policy evaluations for practitioners. It focused on logic model theories and their application. In FY 2014 and FY 2015, the logic model became a theme for the national-level training in Tokyo.

(4) Evaluation exhaustion and evaluation anxiety

Discussions in Japan on the negative aspects of evaluation are common. Recently, evaluation fatigue has been considered a problem of *Gyosei Hyoka* in Japan. Evaluation exhaustion as a negative outcome receives little attention in other countries. Until recently, evaluation anxiety was mostly linked to testing and learning anxiety in pedagogy. Blackmore (2003) is one of the few scholars who was interested in evaluation fatigue as a broad academic topic. The OECD (1997, p. 23) has referred to “evaluation exhaustion” in the context of scientific research evaluation.

The concepts are important because the behavioral and psychological results of evaluation anxiety among officials under evaluation create psychological and temporal burdens for the entire municipality, including evaluators and the central divisions of evaluation activities, especially after the dissemination of the evaluation systems when the original meanings of evaluations become unclear. After the booms of the introduction, the burden on officials was focused among the Japanese municipalities. Related to the concept, the adverse effects of the excessive evaluation anxiety were discussed in the *American Journal of Evaluation* (Donaldson et al., 2002).

Evaluation anxiety entered Western scholarship in about 2000 (e.g., Bechar & Mero-Jaffe, 2014; Donaldson et al., 2002; Taut & Brauns, 2003). Donaldson et al. (2002) described the nature of evaluation anxiety and the excessive evaluation anxiety being addressed in the United States and Europe, concluding that “evaluation anxiety refers to the set of (primarily) affective, and also cognitive and behavioral responses that accompany concern over possible negative consequences contingent upon performance in an evaluative situation” (p. 262). According to these scholars, excessive evaluation anxiety is a response to an excessively anxious situation.

The following analysis of surveys does not deal directly with these concepts. However, it discusses how the

significance of government evaluations in Japan has been challenged since the initial introduction of the systems by the local government more than 20 years ago and how each aspect of the systems has changed.

(5) Methods

1 Nationwide surveys of Japanese Cities

This study analyzed data on Japanese municipalities derived from surveys conducted by the author and Hiroshima University in 2006, 2014, and 2015. The 2006 and 2014 surveys (Nationwide Survey of Japanese Cities Regarding Government Evaluations) were on city municipalities in Japan, and the 2015 survey investigated town and village municipalities (Nationwide Survey of Japanese Towns and Villages).

Nationwide Survey of Japanese Cities Regarding Government Evaluations, 2006

In 2006, the survey of cities was conducted in October. As of January 2006, there were 764 cities in Japan, excluding the ordinance-designated cities such as Osaka, Nagoya, and others. The targets of the survey were 764 city officials. The response rate was 71.1% ($n = 543$). A summary of the 2006 survey results was published in the *Regional Economic Studies* of Hiroshima University (Ito, 2007). The following tables referring to the 2006 survey use the data of Ito (2007).

Nationwide Survey of Japanese Cities Regarding Government Evaluations, 2014

In 2014, the author conducted the survey of cities in November and December. As of January 2014, there were 790 cities in Japan, including ordinance-designated cities. The targets of the survey were 790 city officials. The response rate was 65.1% ($n = 514$). A summary of the 2014 survey results was published in a Japanese journal (Moteki, 2015).

2 Nationwide Survey of Japanese Towns and Villages

Data on smaller municipalities were collected via mail in 2015. In November, questionnaires were mailed to all legally designated small municipalities (745 towns and

183 villages as of April 1, 2015) asking about their government evaluation practices. The questionnaire comprised 26 closed-response questions, some of which had follow-up questions, and 4 open-ended questions. By the deadline (December 25, 2015), 396 valid questionnaires had been returned (response rate of 42.7%).

(6) Results

1 Implementation of government evaluations

Tables 1 and 2 show the population distributions of municipalities. The towns and villages were most likely to have 10,000 to 19,999 residents ($n = 129$, 32.7%), followed by those with 1,000 to 4,999 ($n = 87$, 22.1%). The cities were most likely to have 30,000 to 49,999 residents ($n = 113$, 22.0%), which was closely followed by those with 100,000 to 199,999 ($n = 108$, 21.1%) and those with 50,000 to 69,999 ($n = 107$, 20.9%) residents.

Table 3 shows that 84.6% of the cities were engaged in implementing *Gyosei Hyoka*. About 37.4% of the smaller municipalities were engaged in conducting *Gyosei Hyoka* and about half (49.2%) of them were not, which was markedly different from the situation among the cities. Figures 1 and 2 show the years in which the municipalities introduced *Gyosei Hyoka*. The most common period was 2004 through 2007 for cities and

Table 1 Population Distribution of Cities in the Nationwide Survey of Japanese Cities on Government Evaluations, 2014

Level	Count	Ratio
Less than 30,000	43	8.4%
30,000 to 49,999	113	22.0%
50,000 to 69,999	107	20.9%
70,000 to 99,999	72	14.0%
100,000 to 199,999	108	21.1%
200,000 to 299,999	22	4.3%
300,000 to 499,999	29	5.7%
500,000 and more	19	3.7%
Total	513	100.0%

Table 2 Population Distribution of Cities in Nationwide Survey of Japanese Towns and Villages on Government Evaluations, 2015

Level	Count	Ratio
Less than 1,000	16	4.1%
1,000 to 4,999	87	22.1%
5,000 to 6,999	43	10.9%
7,000 to 9,999	57	14.5%
10,000 to 19,999	129	32.7%
20,000 to 29,999	34	8.6%
30,000 to 39,999	20	5.1%
40,000 and more	8	2.0%
Total	394	100.0%

Table 3 Implementation of Government Evaluations (Cities and Small Municipalities)

	Cities (2014)		Towns/Villages (2015)	
	Number of cases	Percentage	Number of cases	Percentage
Implementing	435	84.6	148	37.4
Planning to implement	17	3.3	29	7.3
Not implementing	21	4.1	195	49.2
Suspended or abolished	41	8.0	24	6.1
Total	514	100.0	396	100.0

smaller municipalities. A comparison of the two distributions suggests that implementation in towns and villages occurred a little later than they did in cities.

Table 4 presents the types of evaluations of cities in 2014 and smaller municipalities in 2015. The *Jimujigyo Hyoka* ($n = 356$, 81.8%), Citizen Participation and Citizen Needs Evaluation ($n = 80$, 18.4%), Cost Management Method ($n = 52$, 12.0%), and Benchmarking and Policy Indicators ($n = 43$, 9.9%) were the most common types. Multiple answers were allowed. The *Jimujigyo Hyoka* was implemented in 356 cities, representing 81.8% of the cities that were engaged in implementing *Gyosei Hyoka*. *Kaizen* Program Total

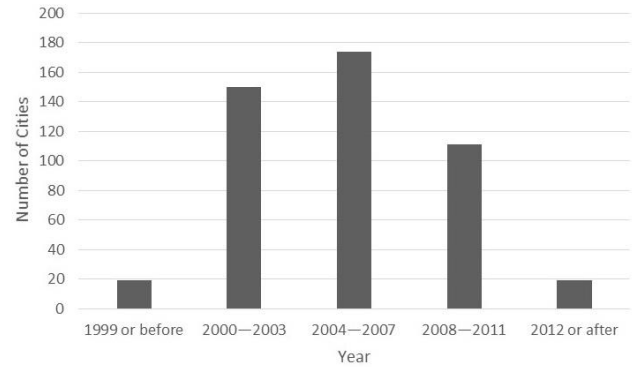


Figure 1 Introduction of Government Evaluations in Cities (Nationwide Survey of Japanese Cities on Government Evaluations, 2014)

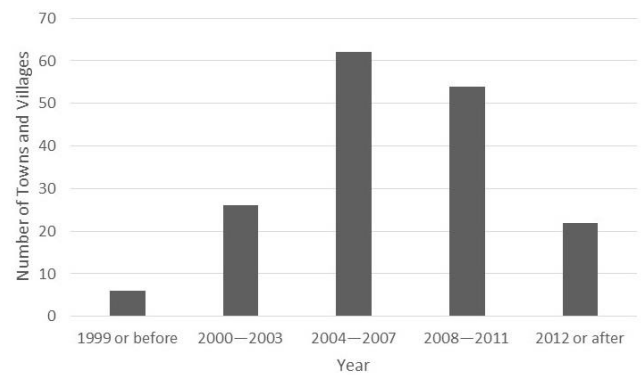


Figure 2 Introduction of Government Evaluations in Towns and Villages (Nationwide Survey of Japanese Towns and Villages on Government Evaluations, 2015)

Quality Management Method, also known as Quality Circle, is an effort to examine and discuss ways to improve operations at Japanese automobile manufacturing facilities. The *Kaizen* Total Quality Management Method was implemented in 17 cities, representing 3.9% of the cities that were engaged in implementing *Gyosei Hyoka* ($n = 6$, 4.1% for towns and villages). The *Kaizen* Total Quality Management Method itself is implemented less than other methods.

The concept of *Kaizen* is introduced as *Hoshin Kanri* (policy deployment) in the United States as a Japanese-style method of business improvement (Akao,

Table 4 *Types of Government Evaluations in Cities (2014) and Towns/Villages (2015)*

Cities with Implemented Evaluation Systems (2014; n=435)^a		
Type of system	Number of cases	Percentage of implementing cities
<i>Jimujigyo Hyoka</i> : Cross-sectional program evaluation system	356	81.8
Evaluation of public construction program by municipalities	5	1.1
Benchmarking and policy indicators	43	9.9
Citizen participation and citizen needs evaluation	80	18.4
<i>Kaizen</i> Program Total Quality Management Method	17	3.9
Cost Management Method	52	12.0
Logic models	13	3.0
Other	28	6.4
Towns and Villages with Implemented Evaluation Systems (2015; n=148)^a		
Type of system	Number of cases	Percentage of implementing municipalities
<i>Jimujigyo Hyoka</i> : Cross-sectional program evaluation system	140	94.6
Evaluation of public construction program by municipalities	4	2.7
Benchmarking and policy indicators	13	8.8
Citizen participation and citizen needs evaluation	22	14.9
<i>Kaizen</i> Program Total Quality Management Method	6	4.1
Cost Management Method	12	8.1
Logic models	2	1.4
Other	13	8.8

^aMultiple responses were accepted.

1991). *Kaizen*-related terms such as Management Cycle and Plan Do Check Action (PDCA) Cycle are often mentioned for introducing *Gyosei Hyoka*, including *Jimujigyo Hyoka*. National training in policy evaluations by the Ministry of Internal Affairs and Communications held in FY2019 stressed the importance of the Management Cycle, including the PDCA cycle as functions of *Gyosei Hyoka* in slide materials for E-Learning, Chapter 1 (https://www.soumu.go.jp/main_sosiki/hyouka/seisaku_n/seisaku_forum.html). Some practitioners and researchers in Japan seem to believe that such a PDCA cycle originated in Europe and the United States because these words seem to be English and are abbreviations of alphabets. However, these concepts originated from the *Kaizen* movement in Japan, and the evaluation community in Europe and the United States rarely mention the importance of evaluation by referring to the Management Cycle or PDCA. Yamaya (2016) stated that PDCA, unfamiliar to overseas evaluation practitioners and scholars, has roots in the Japanese-style QC activities *Kaizen* at the production site, emphasized in corporate management.

Jimujigyo Hyoka, the standard evaluation method, is usually intended to be a side-by-side cross-sectional *ex post* (evidence-based) evaluation of all municipal organizations and programs. These *ex post* evaluations use a performance measurement method that differs from program evaluation concepts, which mainly takes an *ex ante* (forecasting) approach to evaluation. The 2015 questionnaire (towns and villages) asked the respondents about when they implemented their *Gyosei Hyoka*. The question allowed for multiple responses, and 94.6% ($n = 140$) of the smaller municipalities that had implemented evaluations ($n = 148$) had followed the *ex post* approach, whereas 18.0% ($n = 31$) conducted evaluations during their programs and 25.0% ($n = 43$) implemented evaluations before their programs began.

As for the frequency with which municipalities

conducted evaluations, almost 90% of cities and smaller municipalities conducted annual evaluations (Table 5). About 4.9% of the cities implemented biannual evaluations, compared with only 2.3% of towns and villages, suggesting differences between cities and smaller municipalities in terms of personnel and other resources allocated to evaluation activities.

Table 5 *Frequency of Government Evaluations*

Variable	Cities (2014)		Towns/Villages (2015)	
	Number of cases	Percentage	Number of cases	Percentage
Biannual	23	4.9	4	2.3
Annual	424	89.8	151	87.8
Biennial or triennial	13	2.8	4	2.3
Other	12	2.5	13	7.6
Total	472	100.0	172	100.0

Table 6 indicates the reasons why 248 towns and villages had not implemented evaluations. The respondents had seven options, and they were allowed to choose multiple reasons. The most common reason was the lack of resources for evaluation, which was chosen by 94.9% of the municipalities, followed by the lack of knowledgeable staff, which is also a resource. These results indicate that towns and villages could not easily implement evaluation systems because they lacked the necessary resources to do so.

2 Reasons to abolish or suspend government evaluations

Table 7 presents the reasons why municipalities abolished or suspended government evaluations. The most common reason was that the evaluations involved too much work (43.9% of cities and 24.0% of smaller municipalities).

3 Intentions to enhance government evaluations

The 2006 and 2014 surveys of cities also asked about the

Table 6 *Reasons for Non-implementation of Government Evaluations in Towns and Villages in 2015 (n=248 Municipalities Not Implementing Government Evaluations)^a*

Reason	Number of cases	Percentage
The municipality is busy managing municipal mergers.	3	1.5
No staff with expertise in conducting government evaluations.	128	65.6
Local governments do not have enough resources to allocate the necessary personnel or to establish organizations to conduct evaluations.	185	94.9
The top official (the mayor) does not perceive a need to conduct government evaluations.	2	1.0
Staff members are resistant to government evaluations.	10	5.1
The effect of government evaluation is not clear.	47	24.1
Other:	11	5.6

^a Multiple responses were accepted.

anticipation of a change in the evaluation system. In 2006, the respondents were most likely to expect to enhance their systems, whereas in 2014 they were most likely to expect to maintain their current level of evaluation. In 2015, the respondents from towns and villages were most likely to expect to maintain their current levels and the distribution of responses was similar to that of cities in 2014. Table 8 presents the similarities and differences across time for cities and between cities and smaller municipalities. In 2006, 78.6% of the cities wanted to expand *Gyosei Hyoka*, but in 2014, it was down to 39.9%. As for towns and villages in 2015, 29.5% wanted to expand *Gyosei Hyoka*, which was lower than what it was in the previous year. The largest item for the column was the intention to maintain *Gyosei Hyoka* (58.9%). To consider the reasons for this increase, I examine the result of Q20 ("Do you think that the government evaluation in your city is

working?") in 2006 and 2014. In 2014, 1.7% of respondents answered that "It works pretty well" and 45.0% said "It works reasonably well." In 2006, these percentages were 5.1% and 36.5%. The total percentage of local governments choosing either response has thus declined, suggesting that the role played by the government evaluation system in each city is reducing over time. As a result, the number of local governments wishing to expand the system has decreased.

Table 7 *Reasons for the Abolition or Suspension of Government Evaluations in 2014 (Cities) and 2015 (Towns/Villages)*

Variable	Cities		Towns/Villages	
	Number of cases	Percentage	Number of cases	Percentage
Weakens administrative improvements	8	19.5	4	17.4
Requires too much employee work	18	43.9	5	21.7
Change of mayor	4	9.8	3	13.0
Criticism from the local assembly	0	0.0	0	0.0
Other	11	26.8	11	47.8
Total	41	100.0	23	100.0

Table 8 *Intentions to Enhance Government Evaluations: Cities in 2006 and 2014 and Towns and Villages in 2015*

Variable (Intend to...)	2006 (Cities)		2014 (Cities)		2015 (Towns/Villages)	
	#	%	#	%	#	%
Extend/expand	276	78.6	168	39.9	43	29.5
Maintain	64	18.2	208	49.4	86	58.9
Simplify/downsize	21	6.0	45	10.7	17	11.6
Total	351	100.0	421	100.0	146	100.0

4 Reasons for implementing government evaluations

Table 9 compares the reasons for implementing evaluations between cities and smaller municipalities. "Improving fiscal efficiency" was the most common reason and "Increase employees' knowledge" was the next most common among cities and smaller municipalities. However, the city respondents were more likely than those from smaller municipalities to choose the two items regarding citizens, and the respondents from the smaller municipalities were more likely than those from cities to choose "Improve administrative efficiency."

Table 9 *Reasons for Implementing Government Evaluations (Multiple Responses Allowed)*

Variable	Cities (2014)		Towns/Villages (2015)	
	#	% ^a	#	% ^b
Accountability to citizens	213	49.8	48	32.4
Improvement of citizen satisfaction	80	18.7	22	14.9
Increase employees' knowledge	215	50.2	78	52.7
Improve administrative efficiency	133	31.1	70	47.3
Improve fiscal efficiency	217	50.7	115	77.7
Other	22	5.1	3	2.0
Total	880		336	

^aProportion of cities implementing government evaluations (n=435)

^bProportion of small municipalities implementing government evaluations (n=148)

About half the city and the town/village respondents indicated that one reason for implementing evaluations was to increase employees' knowledge. A secondary question asked the respondents about the aspects of

employees' knowledge that the evaluations were expected to improve (Table 10). "Finances" and "Setting goals and managing policy progress" were particularly important aspects. Respondents from cities were more interested than those from smaller municipalities in terms of citizen satisfaction.

In 2014 and 2015, the respondents were asked about problems with government evaluations. The most commonly reported problems in cities in 2014 were "Results of evaluations are unclear despite the burdensome evaluation activities" and "Quality of contents of evaluation sheets varies across departments and officials in charge," which were indicated by about two-thirds of the city respondents (Table 11, numbers 3 and 9). "Evaluation activities rather than the results are becoming the goal" was considered a problem for more than half the city respondents (Table 11, number 1). The most common problems among the 2015 respondents in towns and villages also were numbers 1, 3, and 9 in Table 11. The percentage of these three items for towns and villages are less than that of

Table 10 *Reasons to Increase Employees' Knowledge*

Variable	Cities (2014)		Towns/Villages (2015)	
	#	% ^a	#	% ^b
Manage finances	170	79.1	64	82.1
Improve citizens' satisfaction	129	60.0	37	47.4
Improve productivity/ efficiency	138	64.2	41	52.6
Set goals and manage policy progress	199	92.6	66	84.6
Improve administrative efficiency	4	1.9	2	2.6
Total	640		210	

^aProportion of cities implementing government evaluations (n=435)

^bProportion of small municipalities implementing government evaluations (n= 148)

Table 11 *Problems with Government Evaluations Reported by Cities in 2014 (n=514) and Towns and Villages in 2015 (n=396)*

No	Problem	Cities		Towns/Villages	
		n	%	n	%
1	Government evaluation activities rather than the results are becoming the goal.	233	55.7	73	42.4
2	Government evaluation activities are not well known to the public, and there is little response when they are released.	120	28.7	27	15.7
3	Results of government evaluations are unclear despite the burdensome evaluation activities.	277	66.3	86	50.0
4	As it is an internal assessment, it tends to be used to justify budgets and organizational activities.	35	8.4	17	9.9
5	Uniformly dealing with areas suitable for evaluation and areas not suitable for evaluation.	188	45.0	53	30.8
6	It is impossible to compare and evaluate different fields using a uniform method.	89	21.3	30	17.4
7	The mission and purpose of government evaluation is unclear.	56	13.4	14	8.1
8	A government evaluation method is not established.	62	14.8	34	19.8
9	The quality of the contents of evaluation sheets varies across departments and officials in charge.	264	63.2	93	54.1
10	Government evaluation is not relied on for budgeting, reviewing programs, and so on.	114	27.3	18	10.5
11	It is difficult to evaluate programs annually because there are programs that do not produce results within a single year.	25	6.0	11	6.4
12	There is a time lag between organizing government evaluations and budgeting process.	80	19.1	36	20.9
13	As the relationship between government evaluations and administrative oversight performed by local councils is not clear, problems have arisen.	12	2.9	5	2.9
14	Budget changes alone are emphasized, and programs that should be enhanced are ignored.	26	6.2	10	5.8
15	Other:	24	5.7	5	2.9

the cities in 2014. Here, two-sample tests for comparing proportions for number 3, “Results of evaluations are unclear despite the burdensome evaluation activities,” are conducted. The population ratio of the cities is greater than that of towns and villages, statistically significant at the 1% level. The differences between the city and smaller municipalities suggest that the public and the applicability to specific programs were relatively less important problems and applicability to internal finances and lack of a standard method were relatively more important problems in the smaller municipalities.

(7) Conclusion

This study revealed the characteristics and problems of *Gyosei Hyoka* in Japanese municipalities. The dominant method was *Jimujigyo Hyoka* (Cross-sectional Program Evaluation System), and more than 80% of the municipalities that had implemented evaluations were using *Jimujigyo Hyoka*. Cities tended to implement evaluations first, and the survey data suggested differences between cities and smaller municipalities in terms of implementation and challenges. Some of the differences found were with respect to the implementation rates, reasons for implementation, and the types of problems they were encountering. Thus, smaller municipalities need different approaches towards evaluation. Instead of *Jimujigyo Hyoka*, which significantly demands work requiring specialized human resources, the logic model method may be effective in smaller municipalities for evaluating outcomes and results because it needs fewer resources to accomplish evaluation activities. The logic model method differs from *Jimujigyo Hyoka* in that it does not cover all the programs in principle, but rather concentrates on a specific program. This method was incorporated into the national training in policy evaluation by the MIC from 2013 to 2015. Officers from town and village governments are included in the training and an increasing number of town and village

employees understand the method. In addition, evidence-based policymaking systems, which are attracting attention in Japan, sometimes incorporate the logic model method as a component (as is the case in Hiroshima Prefecture).

One important finding was that the smaller municipalities were relatively less interested in conducting government evaluations because of their limited resources. Fiscal constraints tend to limit the number of public officials in small municipalities, and those officials tend to lack the expertise needed to conduct evaluations. Thus, officials in relatively small municipalities need focused training to obtain the skills necessary to implement evaluations.

One of the major limitations of this study is that it took four years to write this paper in English after the survey conducted in 2015 and the conference presentation in the United States in 2016, partly because of the author's intra-university transfer. Hence, this study was unable to capture the changes in these five years. The MIC report published in June 2017, which examined the situation in each municipality as of October 1, 2016, is the latest result of the government's research. According to this research, the number of cities and Tokyo's 23 special wards implementing government evaluations increased slightly from 82.8% in 2013 to 83.5% in 2016. By contrast, the percentage of towns and villages increased by four percentage points from 34.9% in 2013 to 38.9% in 2016, but it is still below 40%.

The report of the MIC survey published in 2017 examines why no government evaluation was planned to be introduced. The small size of the local government and its difficulty in organizing staff to implement the evaluation (*Taisei ga Torenai*) was the most common answer (61.0%), followed by that no evaluation method and criteria have been established (31.4%). These results are in line with the findings of this study. In the recent literature, two studies dealing with municipal

government evaluations published after 2017 have included the word restructuring (*Sai-kochiku*) in their titles. Future work should aim to research this topic again by reviewing and replacing some of the question items used in this survey to understand the current trend of municipalities abolishing, reintroducing, and significantly revising government evaluation systems.

Acknowledgements

Funding from an affiliated institution (grants for young researchers, Hiroshima University) supported this study. A version of this article, titled "Government Evaluations: Tailoring Evaluations for Public Organizations and Evaluation Fatigue," was previously presented at the American Evaluation Association annual meeting, "Evaluation 2016" in Atlanta, GA, on October 28, 2016. I am grateful to the session chair and participants for their helpful comments.

References

- Akao, Y. (1991). *Hoshin Kanri: Policy deployment for successful TQM*. Great Barrington, MA: Steiner Books.
- Arthur, A., Rydland, L. T., & Amundsen, K. (2012). The user perspective in performance auditing: A case study of Norway. *American Journal of Evaluation, 33*(1), 44–59. <http://doi.org/10.1177/1098214011408283>
- Azuma, N. (2002). Problems and prospects of policy evaluation system of central government in Japan. *Government Accounting Review, 9*, 55–72. Retrieved from http://report.jbaudit.go.jp/english_exchange/volume9/e09d05.pdf
- Bechar, S., & Mero-Jaffe, I. (2014). Who is afraid of evaluation? Ethics in evaluation research as a way to cope with excessive evaluation anxiety: Insights from a case study. *American Journal of Evaluation, 33*(3), 364–376. <http://doi.org/10.1177/1098214013512555>
- Behn, R. D. (2003). Why measure performance? Different purposes require different measures. *Public Administration Review, 63*(5), 586–606. <http://doi.org/10.1111/1540-6210.00322>
- Blackmore, J. (2003). What do students want and how do they know when they get it? Performative universities, quality teaching and academic pedagogies. Paper presented at NZARE at the AARE Joint Conference, December 1–4, Auckland University.
- Donaldson, S. I., Gooler, L. E., & Scriven, M. (2002). Strategies for managing evaluation anxiety: Toward a psychology of program evaluation. *American Journal of Evaluation, 23*(3), 261–273. [http://doi.org/10.1016/S1098-2140\(02\)00209-6](http://doi.org/10.1016/S1098-2140(02)00209-6)
- Hatry, H. P. (2013). Sorting the relationships among performance measurement, program evaluation, and performance management. *New Directions for Evaluation, 2013*(137), 19–32. <http://doi.org/10.1002/ev.20043>
- Ito, T. (2007). Gyosei Hyoka no "meta hyoka" ni kansuru chosa kekka: Zenkoku 764 toshi ni taisuru jittai chosa (Survey results on "meta-evaluation" of government evaluations of 764 cities nationwide). *Chiiki Keizai Kenkyu (Regional Economic Studies), 18*, 77–93. <http://doi.org/10.15027/28824>
- Johnsen, A. (1999). Implementation mode and local government performance measurement: A Norwegian experience. *Financial Accountability & Management, 15*(1), 41–66. <http://doi.org/10.1111/1468-0408.t01-1-00073>
- Koike, O., Hori, M., & Kabashima, H. (2007). The Japanese government reform of 2001 and policy evaluation system: Efforts, results and limitations. *Ritsumeikan Law Review, 24*, 1–12. Retrieved from <https://www.researchgate.net/publication/237701627>
- Kudo, H. (2015). Bridging the gap between reform practices and literatures: New public service delivery of Japanese local governments. *International Public Management Review, 16*(1), 147–164. Retrieved from <http://journals.sfu.ca/ipmr/index.php/ipmr/article/view/250>
- Lahey, R., & Nielsen, S. B. (2013). Rethinking the

- relationship among monitoring, evaluation, and results-based management: Observations from Canada. *New Directions for Evaluation*, 137, 45–56. <https://doi.org/10.1002/ev.20045>
- Ministry of Internal Affairs and Communications. (2017). *Chiho kokyo dantai ni okeru Gyosei Hyoka no torikumi jokyō (Status of government evaluation efforts by local governments)*. Retrieved from https://www.soumu.go.jp/fiken/02gyosei/04_04000062.html.
- Morisugi, H. (2000). Evaluation methodologies of transportation projects in Japan. *Transport Policy*, 7(1), 35–40. [http://doi.org/10.1016/S0967-070X\(00\)00013-5](http://doi.org/10.1016/S0967-070X(00)00013-5)
- Moteki, Y. (2015). Gyosei hyoka bumo go no jichitai ni okeru seisaku hyoka seido no henka to kadai: "Hyoka fuan" gainen wo tegakari ni shita zenkoku chosa ni motodzuku "hyoka dsukare" no kento. (Changes and challenges after the evaluation boom years in Japanese local governments: An examination of evaluation exhaustion based on the data from a nationwide survey focusing on the concept of evaluation anxiety). *Chihoujichi Kenkyu (Journal of Urban Management and Local Government Research)*, 30(2), 1–17.
- Noutomi, I., & Nakanishi, M. (2007). Net results of the Japanese NPM movement at local governments since the mid-1990s: Performance budgeting, total quality management and target based budgeting. *International Journal of Public Administration*, 30(12–14), 1393–1433. <http://doi.org/10.1080/01900690701229574>
- Organisation for Economic Cooperation and Development. (1997). The evaluation of scientific research: Selected experience. OCDE/GD(97)194. Retrieved from [http://oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=OCDE/GD\(97\)194&docLanguage=En](http://oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=OCDE/GD(97)194&docLanguage=En).
- Osborne, D., & Gaebler, T. (1992). *Reinventing government: How the entrepreneurial spirit is transforming the public sector*. Reading, MA: Addison-Wesley.
- Sato, T. (2013a). *Gyosei keiei shisutemu ni okeru Gyosei Hyoka no kozo to kino*. (Structure and function of government evaluations in administrative management system). *Takasaki Keizai Daigaku ronshu (Bulletin of Takasaki City University of Economics)*, 56(2), 53–66. Retrieved from <http://id.nii.ac.jp/1496/00000703/>.
- (2013b). *Hyoka jinzai no ikusei to sono kadai: Jichitai gyosei no hyoka wo megutte* (Development of evaluation human resources and their issues). *Hyoka Quartely (Evaluation Quartely)*, (27), 3–19.
- Sawada, S. (2010). Todofuken ni okeru Gyosei Hyoka no genjo to kadai (Current status and issues of government evaluations in prefectures). *Keiei Senryaku Kenkyu (Management Strategy Review)*, (4), 99–110. Retrieved from https://kwansei.ac.jp/fiba/assets/pdf/journal/studies_in_Banda_2010_p99-110.pdf
- Tabuchi, Y. (2010). Chihou jichitai ni okeru Gyosei Hyoka 12 nen no ayumi to kongo no kadai (Progress of public sector evaluations in local governments in 12 years and future prospects). *Mitsubishi Sogo Kenkyujo Hou (Mitsubishi Research Institute Report)*, (53), 30–53.
- Tanaka, M. (1989). Japanese-style evaluation systems for R&D projects: The MITI experience. *Research Policy*, 18(6), 361–378. [http://doi.org/10.1016/0048-7333\(89\)90023-1](http://doi.org/10.1016/0048-7333(89)90023-1)
- Taut, S., & Brauns, D. (2003). Resistance to evaluation: A psychological perspective. *Evaluation*, 9(3), 247–264. <http://doi.org/10.1177/13563890030093002>
- Uchida, M. (2009). NPM-gata Gyosei Hyoka ni honro sareru shokibo jichitai: Genjo no ninshiki to kongo no tenbo. (Small-scale local governments at the mercy of NPM-type government evaluations: Recognition of the current situation and future prospects) *Nihon Hyoka Kenkyu (Japanese Journal of Evaluation Studies)*, 9(3), 69–82. <https://doi.org/10.11278/jjoes.9.3.69>
- Ueyama, S. (1998). "Gyosei Hyoka" no jidai: Keiei to kokyaku no shiten kara (Era of "Government Evaluation": Perspectives from management and

- customers*). Tokyo: NTT Publishing.
- Yamamoto, K., & Watanabe, M. (1989). Performance auditing in the central government of Japan. *Financial Accountability & Management*, 5(4), 199–217. <http://doi.org/10.1111/j.1468-0408.1989.tb00090.x>
- Yamaya, K. (2016). Jichitai kaikaku to hyoka (Local governments reform and evaluations). *Nihon Hyoka Kenkyu (Japanese Journal of Evaluation Studies)*, 16(1), 1–2. http://evaluationjp.org/files/Vol16_No1.pdf
- (2002). The art of policy evaluation in Japan. In J. E. Furubo, R. C. Rist, & R. Sandahl (Eds.), *International atlas of evaluation* (pp. 337–355). Piscataway, NJ: Transaction Publishers.
- Yokoyama, M. (2009). Evaluation systems in local government: Are evaluation systems useful tools of administrative reform? *Interdisciplinary Information Sciences*, 15(2), 243–250. <http://doi.org/10.4036/iis.2009.243>
- Yonezawa, A. (2007). Stability amidst a storm of evaluation: Policy trends and practice in higher education evaluation in Japan. In C. Alessandro (Ed.), *Quality assessment for higher education in Europe* (pp. 95–103). Portland, ME: Portland Press.

Note: The English titles in parentheses are the author's translation of the original Japanese titles.

(Received: July 3, 2020)

(Accepted: October 17, 2020)