Methods of Program Evaluation and Implementation of Comprehensive Evaluations

Nobuo AZUMA*
(Director, Study Division, Board of Audit)

I. Introduction

The central-government reform carried out in Japan in January 2001 prompted the introduction of a policy evaluation system in the ministries and agencies, and in April 2002 the Government Policy Evaluation Act was put into force to raise the effectiveness of this system. Grounded on evaluations conducted by ministries/agencies on the policies for which they are respectively responsible, this policy evaluation system allows ministries/agencies to selectively use the project evaluation method, the performance measurement method, and the comprehensive evaluation method. The last of these evaluation methods[1] explores and analyzes from various angles the manifestation of policy effects for a specified topic, ascertains problems with policies, and analyzes their causes. The comprehensive evaluation method was the first evaluation method introduced into Japan as a system, and some ministries began in FY2002 to prepare evaluation reports in accordance with the Government Policy Evaluation Act.

In the US, program evaluation has been pursued since the 1960s, primarily in federal departments and agencies. Program evaluation is an evaluation method that systematically analyzes policy effects utilizing social sciences techniques, and appears to correspond to Japan's comprehensive evaluation. The US mandates that federal departments/agencies conduct program evaluations utilizing the randomized experimental model or other analysis techniques, and evaluation techniques and quantitative analysis techniques have thus been developed in accordance with systematic evaluation theory and adopted in actual practice.

Given the need to make the manifestation of policy effects explicitly clear in comprehensive evaluations, program evaluation techniques are regarded as useful tools by Japan's ministries/agencies in carrying out comprehensive evaluations. This article introduces program evaluation techniques and matches up the comprehensive evaluations implemented by ministries/agencies in FY2002 and FY2003 with program evaluation techniques (all the opinions expressed in this article are those of the author, and do not necessarily reflect the official position of the Board of Audit to which the author belongs).

---

* Born in 1956. Graduated from Yokohama National University with a bachelor's degree in economics in 1980. Graduated from University of Rochester in the U. S. with a degree of MBA in 1986. Joined the Board of Audit of Japan in 1980 and followed his career as Assistant Chief, the Second Education Audit Division, Assistant Chief, Trade and Industry Audit Division, and Senior Accounts Verification Officer, Audit Division of Finance. Worked at Consulate General of Japan in New York from 1990 through 1993. Presently serves as Director of the Study Division, Board of Audit. Also lectured in economics at Nagoya University in 2003.

1) "Basic Guidelines on Policy Evaluation (approved by the Cabinet on December 28, 2001)" (Appendix) [Comprehensive evaluation method].

II. Program evaluation techniques[2)]

1. Theoretical background to program evaluation

In order to understand program evaluation techniques, it is essential to know their theoretical background. The policies of ministries/agencies on program evaluation appear to be as follows.

(1) Policy system

Ministries/agencies implement policies to resolve issues arising in people's lives and in the society/economy and, examined from the perspective of goals and means, the policy system comprises "policies (narrowly defined) → programs → projects." Here, (i) "policies (narrowly defined)" denote the basic objectives for resolving a specific administrative issue, (ii) "programs" are the concrete objectives for achieving the policies (narrowly defined) and are the sum of organizational government activities, and (iii) "projects" are individual government activities that constitute the specific policy means for achieving the programs. In many instances, the policies of the ministries/agencies feature multiple programs for each policy (narrowly defined) and multiple projects for each program, resulting in a three-tier pyramid structure. Thus policies (narrowly defined), programs and projects are interconnected as goals and means while on the whole comprising a single policy system.

(2) Theory

The policies of the ministries/agencies have been designed in accordance with a theory marked by a process consistent through the achievement of the policies' objectives. This theory is a "supposition" linking causes and results in a chain: should cause 1 give rise to result 1, then this result 1 becomes cause 2 and produces result 2, which in turn becomes cause 3 and leads to result 3. The process through the achievement of the policy objectives is designed on the basis of this chain-link "supposition" and the failure of any link in this supposition to function properly will cause a break in the policy and prevent it from achieving the improvement effects it seeks.

(3) Logic model

The policies of ministries/agencies, regarded in terms of cause-and-result relationships in line with the theory, have an impact on people's lives and on the society/economy through the processes of "input → activities → output → (external factor) → outcomes." This flow is termed a logic model in program evaluation. Here, (i) "input" is the input of resources necessary to implement government activities, (ii) "activities" are organizational government activities to deliver output, (iii) "output" is the government services delivered to resolve an issue arising in people's lives or in the society/economy, (iv) "external factors" are factors other than this output that have an impact on the outcome, and (v) "outcomes" are the improvement effects on people's lives and the society/economy; there are individual outcomes (lower outcomes) and comprehensive outcomes (higher outcomes). Thus input, activities, output and outcomes are interconnected as causes and results while on the whole comprising a single logic model.

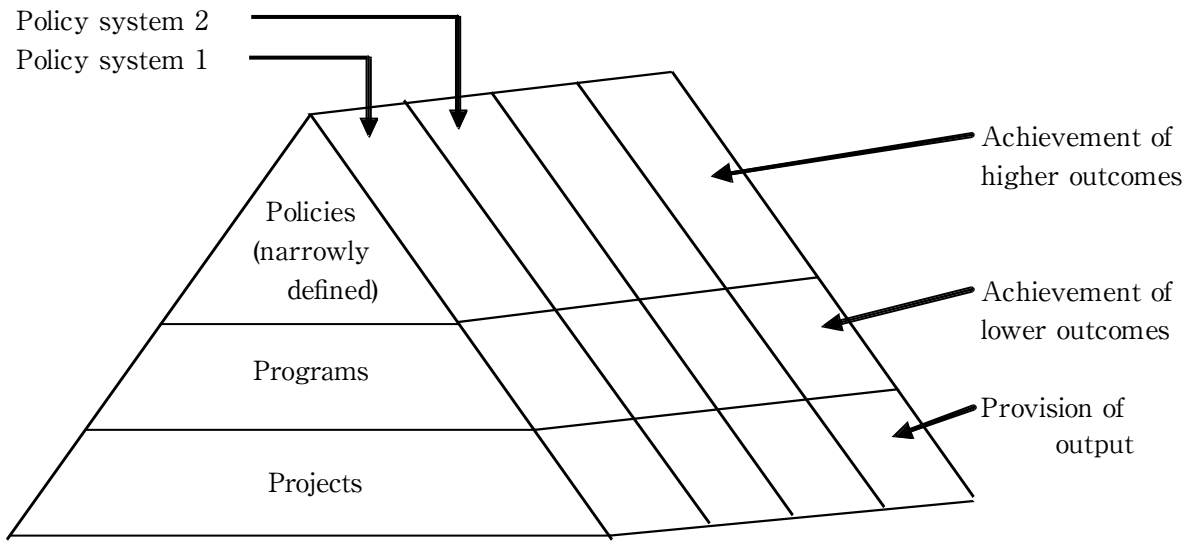
(4) Policy system and logic model

Viewing the policies of ministries/agencies in the context of the relationship between the policy system and the logic model, projects are individual government activities and the means for achieving programs, while output corresponds to the government services delivered by projects and lower outcomes to the improvement effects set as objectives for the programs. When there are multiple projects for a single program, multiple outputs will be provided to achieve the lower outcomes.

2) Section II owes a great deal to RYU Yoshiaki and SASAKI Ryo (2004).

With programs as the means for achieving policies (narrowly defined), higher outcomes correspond to the improvement effects set as objectives for the policies (narrowly defined). When there are multiple programs for a single policy (narrowly defined), multiple lower outcomes for achieving higher outcomes will have improvement effects on people's lives and the society/economy (see Figure 1 on the relationship between the policy system and the logic model).

Figure 1 Relationship between policy system and logic model



2. Evaluation techniques for program evaluation

When the program[3] does not produce effects, this may be attributed to (i) the program not being designed with consistency, (ii) the program not being implemented as designed, or (iii) the program not using effective policy means. It may also be possible that the program has input costs greater than its outcomes. To clarify such circumstances, the following four evaluation techniques have been adopted in program evaluation.

(1) Theory evaluation

Theory evaluation is a technique for evaluating whether or not the policy systems in ministries/agencies proposing the program have been designed with a clear-cut relationship between the policy objectives and the policy means, and whether or not a consistent logic model has been designed for the chain relation between cause and result. This evaluation technique is based on the idea that outcomes will not be achieved if the logic model has not been designed with consistency; unless the supposition chain from input to outcome functions properly, the program, even if implemented, will fail partway through and not achieve the improvement effects sought. When theory evaluation

3) Program evaluations can also evaluate any tier of policies constituting a policy system, but from II.2 onward the assumption will be that program evaluations focus on the program that uses projects as policy means.

produces a negative evaluation, a review of the logic model is conducted.

(2) Process evaluation

Process evaluation is a technique for evaluating whether or not the planned quantitative/qualitative output is being delivered by ministries/agencies implementing the program as scheduled in accordance with the pre-designed logic model. This evaluation technique focuses on the "input → activities → output" process within the logic model and is based on the idea that, unless the planned quantitative/qualitative output is delivered as scheduled, the outcome will not be achieved even if the logic model is designed in a consistent form. When process evaluation produces a negative evaluation, reviews are conducted on the quality/quantity and timing of input as well as on the methods/contents of government activities.

(3) Impact evaluation

Impact evaluation is a technique for evaluating whether or not the program implemented by ministries/agencies has an improvement effect (impact) on people's lives and the society/economy. This evaluation technique focuses on the "output → (external factor) → outcome" process and is based on the idea that, unless the output is effective as a means of achieving the outcome, then the outcome will not be achieved even if the planned quantitative/qualitative output is delivered at the scheduled timing. When impact evaluation produces a negative evaluation, reviews are conducted on the quality/quantity and timing of the output as well as on the output as a policy means itself.

(4) Cost-efficiency analysis

Cost-efficiency analysis is a technique for evaluating whether or not the program implemented by ministries/agencies has improvement effects on people's lives and the society/economy greater than the resources input. This evaluation technique is based on the idea that a policy means cannot be regarded as cost efficient if the costs are greater than the corresponding impact, even if the output does have an impact on people's lives and the society/economy. When cost-efficiency analysis produces a negative evaluation, reviews are conducted on the quantity/quality and timing of the output as well as on the output itself as a policy means (see Figure 2 on the relationship between the logic model and evaluation techniques).

The respective evaluation items and analysis techniques for each of these evaluation techniques will be discussed in more detail below.

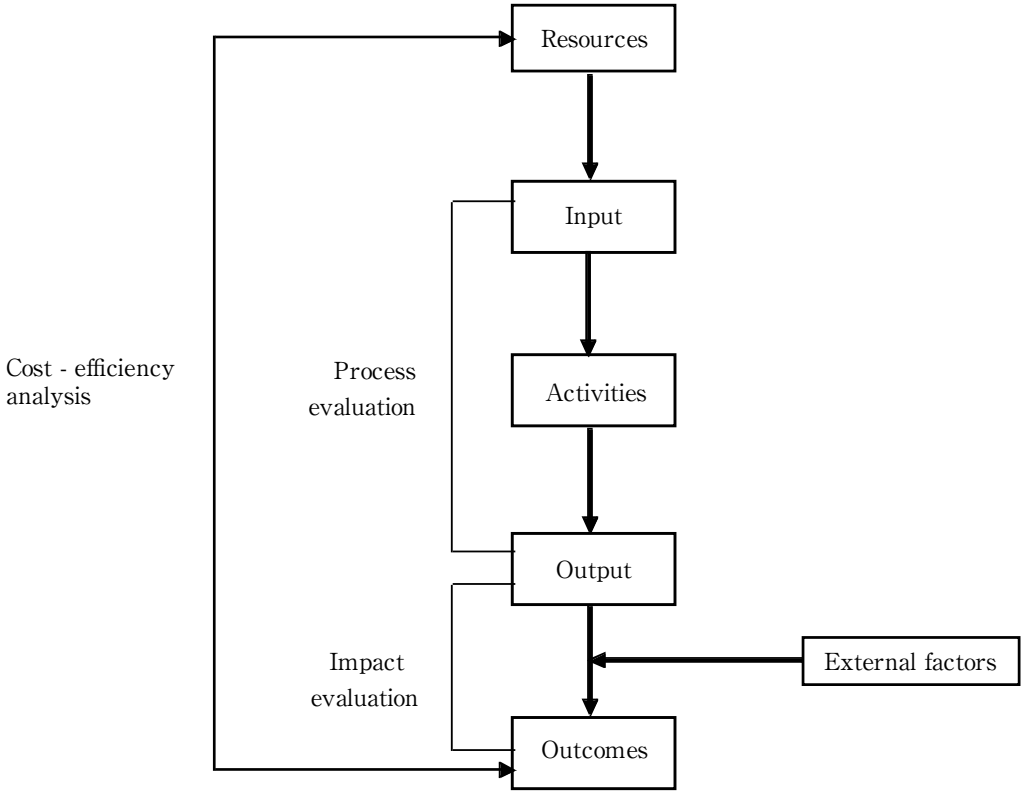
3. Evalution items for theory evaluation

Theory evaluation specifically examines the logic model of "input → activities → output → (external factors) → outcomes" and verifies whether or not a consistent logic model has been designed for the chain relation of causes and results. It is essential in theory evaluation to ascertain the following items on a plan basis and to prepare a flow chart.

(1) Outcomes

Confirmation is made that the policy system clarifies the relationship between policy objectives and policy means, and the specific objectives of the program targeted for evaluation are assessed. To quantitatively measure the success achieved by the program using performance indicators in program evaluation, the program must ultimately be specifically translated into terms by which it can be quantitatively measured. The performance indicator for quantitatively determining the success of the program (outcome performance indicator) is also assessed. This outcome performance indicator is a quantitative evaluation indicator that looks at the changes in people's lives and the society/economy.

Figure 2　Logic model and evaluation techniques



(Note) Theory evaluation specifically examines the logic model itself, and verifies the consistency of the logic model.

(2) Output

An overall picture of projects that serve as the policy means for the program is developed on a plan basis. In addition to the names of the projects that serve as policy means, (i) the implementing entity, (ii) the implementation region, (iii) the implementation period, (iv) the beneficiary qualifications and total number of beneficiaries, and (v) the quantity and quality (output performance indicator) of the output delivered by these projects are examined. Here, "quantity" refers to the scale of output and "quality" to the nature of the output. In a public works project, for example, "quantity" would include such aspects as processing capability and capacity, while "quality" would comprise elements such as standards and structure. The "quantity" of financing services includes the number of loans and the total amount of loans, and the "quality" would encompass the the loan ceiling and loan interest rate for individual loans. An overall picture of these respective projects is developed with respect to the program for which multiple projects serve as the policy means.

(3) Activities

The organizational government activities for delivering output are ascertained on a plan basis. More specifically, the series of operations carried out so that the implementing entity for the projects can deliver output to beneficiaries is examined. In a social security project, for example, these operations would be "collection of contributions → acceptance of applications for benefits → qualification screening → payout of benefits;" in a public works, project they would be "plan → design → estimate of cost → contract → construction → facility operation;" and in financing services, they would be "procurement of financial resources for loans → acceptance of loan applications → loan screening → loan implementation → credit management."

(4) Input

The quantity/quality of resources needed to deliver output is ascertained on a plan basis. Specifically, (i) project costs and other financial resources, (ii) employees and other human resources, (iii) working hours and other temporal resources, (iv) facilities and other physical resources, and (v) beneficiary needs and other informational resources that are input by the implementing entity of projects to deliver output to beneficiaries are examined. Here, "quantity" refers to the scale of the resources and "quality" to the nature of these resources. For financial resources, "quantity" includes the total amount of project costs and "quality" the breakdown of government subsidies as well as copayments, while for human resources "quantity" is the total number of staff and "quality" a breakdown of this staff by position/qualification. It is important to ascertain as comprehensively as possible not just direct project costs and other financial resources but also such indirect project costs as personnel and management costs.

(5) Process theory

Process theory determines the interrelationships among input, activities and output. The output production stage and the output utilization stage are examined separately. At the output production stage, the ways in which resources are combined and the order and timing by which they are input during the "input → activities" process to produce output within the government are studied. At the output utilization stage, the focus is on the timing and the procedure by which beneficiaries use the output within the "activities → output" process. It is important to draw a clear distinction between the production and utilization stages of output because output may not necessarily be used by beneficiaries even if produced within the government.

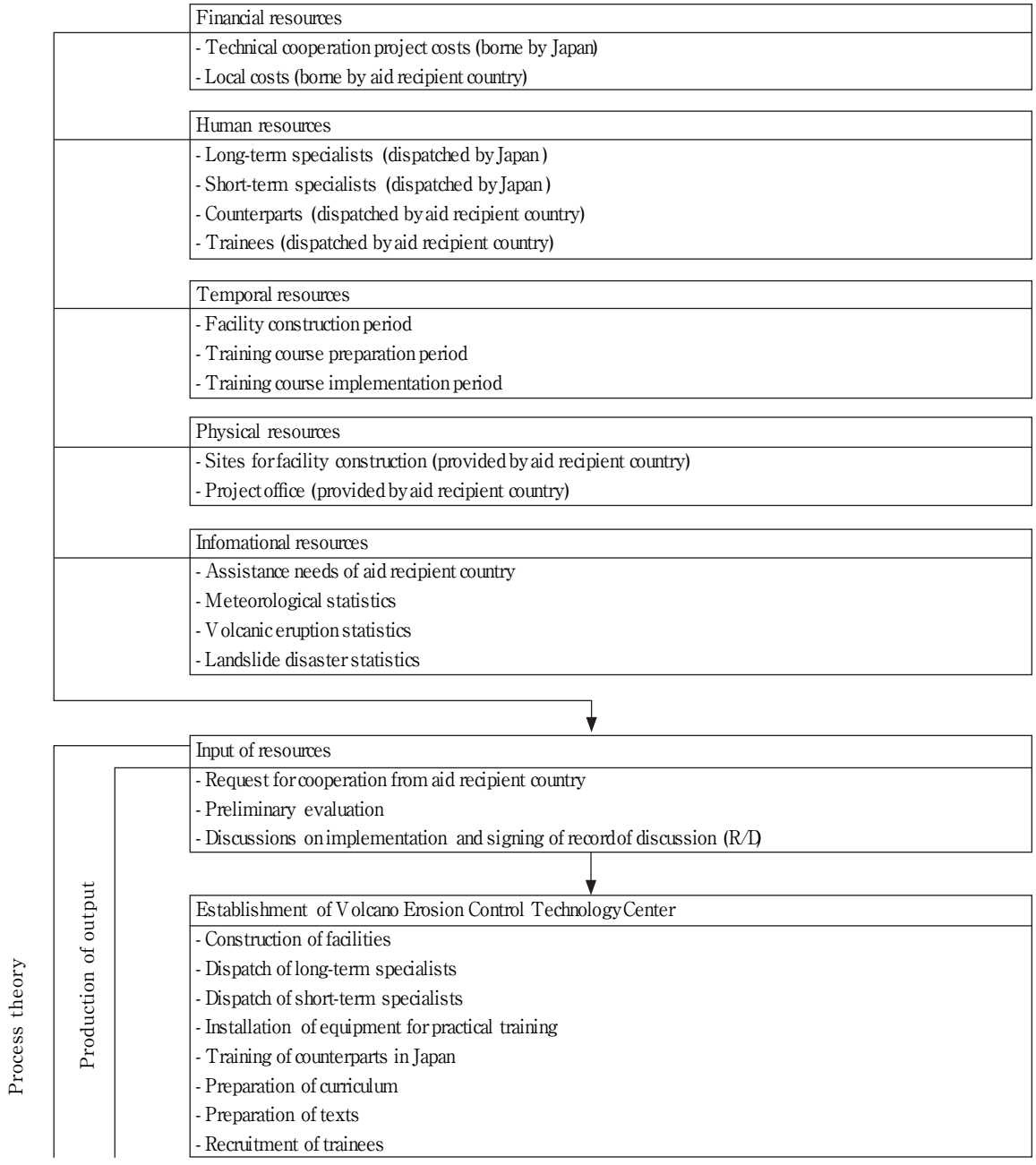
(6) Impact theory

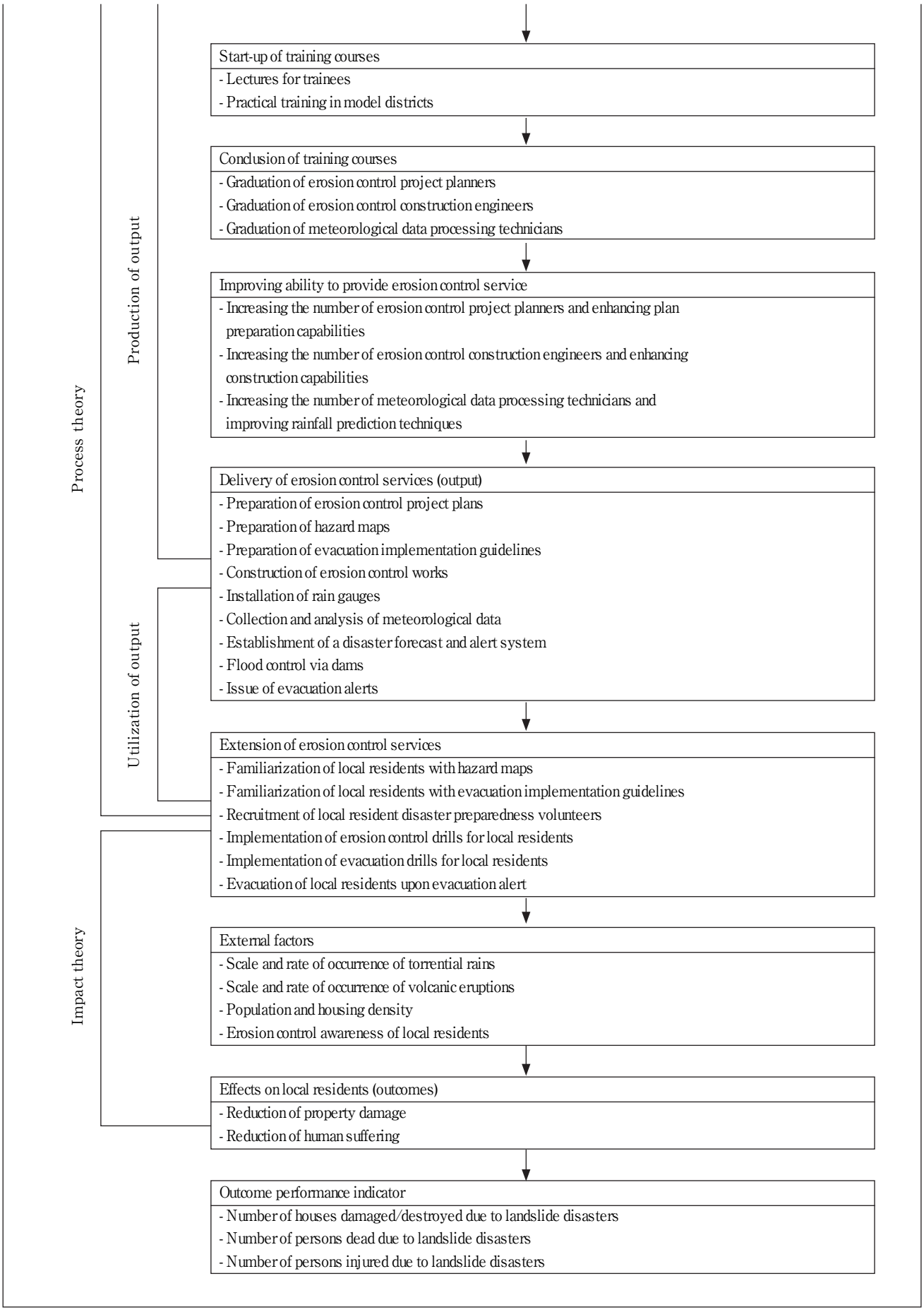
Impact theory ascertains the interrelationship between output and outcome. Specifically, the course by which the output delivered within the "output → (external factors) → outcome" process has an improvement effect on people's lives and the society/economy is examined. Also considered are the external factors anticipated to have an impact on improvement effects (see Figure 3 for a hypothetical example of a logic model presented in flowchart form).

Figure 3 Hypothetical example of a logic model

The "Volcanic Region Comprehensive Disaster Preparedness Project (an ODA project) will be presented here as a simplified hypothetical example of theory evaluation. To reduce the damage due to landslide disasters occurring in the volcanic country that is ODA aid recipient, this ODA project envisions Japan transferring erosion control technology to aid recipient country by accepting trainees, dispatching experts and providing equipment. Specifically, once the Water Resources Development Bureau of the Local Infrastructure Ministry overseeing comprehensive disaster preparedness policy in the aid recipient country has established a Volcano Erosion Control Technology Center, the counterpart receiving the technology transfer will hold training courses for the Bureau's personnel to train ① erosion control project planners, ② erosion control construction engineers, and ③ meteorological data processing technicians. Persons completing the training courses will be involved in preparing evacuation implementation guidelines that clearly indicate evacuation routes and evacuation sites in all locales for emergencies, and issue evacuation alerts based on meteorological information collected and analyzed.

In preparing a logic model via theory evaluation for the "Volcanic Region Comprehensive Disaster Preparedness Project," the following steps would be taken.




Process theory
Production of output
Utilization of output
Impact theory
Start-up of training courses
- Lectures for trainees
- Practical training in model districts
Conclusion of training courses
- Graduation of erosion control project planners
- Graduation of erosion control construction engineers
- Graduation of meteorological data processing technicians
Improving ability to provide erosion control service
- Increasing the number of erosion control project planners and enhancing plan preparation capabilities
- Increasing the number of erosion control construction engineers and enhancing construction capabilities
- Increasing the number of meteorological data processing technicians and improving rainfall prediction techniques
Delivery of erosion control services (output)
- Preparation of erosion control project plans
- Preparation of hazard maps
- Preparation of evacuation implementation guidelines
- Construction of erosion control works
- Installation of rain gauges
- Collection and analysis of meteorological data
- Establishment of a disaster forecast and alert system
- Flood control via dams
- Issue of evacuation alerts
Extension of erosion control services
- Familiarization of local residents with hazard maps
- Familiarization of local residents with evacuation implementation guidelines
- Recruitment of local resident disaster preparedness volunteers
- Implementation of erosion control drills for local residents
- Implementation of evacuation drills for local residents
- Evacuation of local residents upon evacuation alert
External factors
- Scale and rate of occurrence of torrential rains
- Scale and rate of occurrence of volcanic eruptions
- Population and housing density
- Erosion control awareness of local residents
Effects on local residents (outcomes)
- Reduction of property damage
- Reduction of human suffering
Outcome performance indicator
- Number of houses damaged/destroyed due to landslide disasters
- Number of persons dead due to landslide disasters
- Number of persons injured due to landslide disasters

4. Evaluation items for process evaluation

Process evaluation ascertains the performance of the "input → activities → output" process in the logic model and verifies whether or not output is being delivered as planned. It is essential in process evaluation to grasp the following items on a performance basis and then compare them with the plan.

(1) Results of resource input

Whether the resources have been input as planned is verified. If the expected quantitative/qualitative resources are not input as scheduled, then the government activities cannot be carried out as planned. In the hypothetical example given in Diagram 3, for example, a Volcanic Erosion Control Technology Center cannot be constructed unless the aid recipient country is able to procure enough funds to cover local costs and to provide sites for facility construction.

(2) Results of government activities

Whether the government activities have been carried out as planned by the implementing entity after the input of resources is verified. If the government activities are not performed as planned, the expected quantitative/qualitative output cannot be produced on schedule. In the hypothetical example given in Diagram 3, for instance, evacuation implementation guidelines that clearly identify the evacuation routes and evacuation sites in each region in the event of a disaster cannot be prepared, even if construction of the Volcanic Erosion Control Technology Center is completed, if the Local Infrastructure Ministry does not recruit trainees for the training courses or if the persons completing the training courses are reassigned to different bureaus.

(3) Results of output production

Whether output has been produced as planned as a result of government activities is verified. If the planned output is not produced within the government, then beneficiaries cannot use the anticipated quantitative/qualitative output as scheduled. In the hypothetical example given in Diagram 3, for example, local residents will not be able to use erosion control services in the event of a disaster if the persons completing training courses at the Volcanic Erosion Control Technology Center do not prepare evacuation implementation guidelines or do not issue evacuation alerts based on meteorological information collected and analyzed.

(4) Results of output utilization

Whether the beneficiaries have used the output as planned after it was produced is verified. If the beneficiaries do not use the anticipated quantitative/qualitative output as scheduled, then there will be no improvement effect on people's lives and the society/economy. In the hypothetical example given in Diagram 3, for example, landslide disasters cannot be reduced – even if evacuation implementation guidelines are prepared and evacuation alerts issued on the basis of meteorological information collected and analyzed – if local residents do not evacuate prior to a disaster accordingly.

5. Analysis techniques for impact evaluation

Impact evaluation analyzes the "output → (external factors) → outcomes" process within the logic model and measures the impact to verify whether or not the program implemented by ministries/agencies has had an improvement effect on people's lives and the society/economy. In impact evaluation, the parent population for which the program may be carried out is divided into one group for which the program is actually implemented (implementation group) and another for which the program is not implemented (comparison group) so that any difference emerging between these two groups can be measured. If no comparison group is available, differences arising in the status of the implementation group before and after implementation are measured. In either case, it is crucial to

eliminate the impact of external factors on these differences. The following analysis techniques are used for measuring these differences (see Figure 4 on analysis techniques for impact evaluation).

Figure 4 Analysis techniques for impact evaluation

Analysis technique		Comparison group used	Method for establishing comparison group	Timing for establishing comparison group	Reliability of analysis results
Beneficiaries judgment		No			Low
Simple before-after comparison model		No			
Panel study		No			
Interrupted time series model		No			
Cross-section model		No			High
Semi-experimental model	Generic control model	Yes	Statistical	After implementation of the program	Low
	Statistically equated model	Yes	Statistical	After implementation of the program	
	Matching model	Yes	Statistical	Before implementation of the program	
	Regression-discontinuity model	Yes	Statistical	Before implementation of the program	
Randomized experimental model		Yes	Random sampling	Before implementation of the program	High

Source: Table prepared by author by making additions/revisions to the table presented on p.51 in the book written by RYU Yoshiaki and SASAKI Ryo (2004)

(1) Beneficiaries judgment

a. Techniques

Beneficiaries judgment is an analysis technique that has beneficiaries in the implementation group, after the program has been implemented, forecast estimates of the outcome performance indicator had the program not been implemented on the basis of their personal recollections and then has them compare estimates with the outturn of the outcome performance indicator; the difference as acknowledged by the beneficiaries is deemed to be the impact of the program. This analysis technique is implemented through questionnaire surveys and beneficiary interviews.

b. Technical limitations

Beneficiaries judgment has been widely adopted because it is simple, but it does not have high reliability as an analysis technique because the memories of beneficiaries are often vague and uncertain. Variations in the methods for selecting survey participants, for devising survey questions, and for replying to the survey can at times lead to greatly different analysis results.

(2) Simple before-after comparison model

a. Techniques

The simple before-after comparison model is an analysis technique that measures the outturns

of the outcome performance indicator for the same implementation group before and after the implementation of the program, and defines the impact of the program to be the difference in outturns between these two points in time. As a general rule measurements are made directly before and directly after the program is implemented, but year-on-year comparisons may be made on occasion in view of seasonal variables.

b. Technical limitations

The simple before-after comparison model has been widely adopted for its simplicity, but it is not very reliable as an analysis technique because it cannot exclude the impact of external factors on the outcome performance indicator. This analysis technique can only be utilized in cases where it can be justifiably assumed that the "before" and "after" outturns of the outcome performance indicator would have remained at the same level had the program not been implemented. This prerequisite is fulfilled when the two measurements are taken at times separated by only a relatively short interval. As the impact of the program will not necessarily be apparent directly after implementation, this model cannot be utilized for the program whose impact is expected to require some time to emerge.

(3) Panel study

a. Techniques

The panel study is an analysis technique that measures the outturns of the outcome performance indicator for a given period before and after the implementation of the program for a fixed identical implementation group; the impact of the program is defined as the difference emerging in outturns. This analysis technique is often adopted in cases where the impact is not expected to appear directly after implementation of the program.

b. Technical limitations

The panel study requires that the outturns of the outcome performance indicator be measured over a given period but, if the number of samples is reduced due to survey participants changing residences or otherwise dropping out of the survey, the analysis results will become less reliable because of a resultant bias in the measurements. This model is also highly labor-intensive and costly because it is premised on knowing the location of all survey participants over the period in question. Furthermore, the impact of external factors on the outcome performance indicator cannot be eliminated because no comparison is made with a group for whom the program is not implemented.

(4) Interrupted time series model

a. Techniques

The interrupted time series model is an analysis technique that measures the outturns of the outcome performance indicator over an extended period before and after the implementation of the program for the same implementation group; the impact of the program is defined as the differences apparent in the trend of the outturns. The trend of the outturns is visually analyzed via tables and graphs and statistically analyzed via regression analysis. This analysis technique can forecast values of the outcome performance indicator by ascertaining the trend and regularity of the outcome performance indicator.

b. Technical limitations

In the interrupted time series model, the outturns of the outcome performance indicator must be measured over an extended period, but variations in the measurement methods (survey items, survey methods, survey participants, etc.) between the measurement times can have an impact on the measurement results and lower the reliability of the analysis results. The impact of external factors on the outcome performance indicator cannot be excluded because no comparison is made with a group for whom the program is not implemented.

(5) Cross-section model

a. Techniques

The cross-section model is an analysis technique that measures the outturns for the output performance indicators and the outcome performance indicators for multiple implementation groups at a point after implementation of the program to ascertain the correlation between them; the degree of correlation defines the impact of the program. The degree of correlation is visually analyzed via tables and graphs and statistically analyzed via regression analysis. Because of the small number of implementation groups in the sample and variations in the populations by area (metropolises, farming villages, etc.), there are times, in cases where data from a large implementation group greatly influence the analysis results, when the correlation is ascertained by converting the outturns of the outcome performance indicator on a population basis (e.g., per 1000 persons).

b. Technical limitations

The cross-section model requires that data be collected from a minimum number of implementation groups (for example, 25 or more) in order to measure the impact using the disparities in quantity and quality of government services actually delivered. This model can also ascertain the correlation between policy means and improvement effects but, because no comparison is made with a group for whom the program are not implemented, no cause-and-result relationship between policy means and improvement effect can be established nor can external factors be excluded from improvement effects.

(6) Generic control model

a. Techniques

The generic control model considers the average value of the parent population potentially targeted by the program (nationwide average value, prefecture-wide average value, citywide average value, etc.) to be the comparison group. This parent population thus comprises a group for whom the program was implemented and a group for whom the program was not implemented. Therefore, this is an analysis technique that measures the outturns of the outcome performance indicator for the implementation group and the parent population over a given period (or at a particular point in time) following implementation of the program, and that defines the impact of the program as the difference that appears between the outturn of the implementation group and the average value of the parent population. This analysis technique presumes that the parent population is subject to the influence of external factors and that the implementation group is subject to the same influence.

b. Technical limitations

The generic control model can only be adopted in government areas in which such general indicators as nationwide and prefecture-wide averages are available. However, this analysis technique is not highly reliable because no detailed analysis is conducted on how closely external factors affecting the parent population as a comparison group resemble those affecting the implementation group. Furthermore, the reliability of analysis results tends to diminish as the program is implemented over a broader scope because the greater the share of the implementation group in the parent population, the smaller the difference between the outturns of the implementation group and the average value for the parent population.

(7) Statistically equated model

a. Techniques

To establish an implementation group and a comparison group under the statistically equated model, segments sharing similar conditions are statistically extracted after the implementation of the program from the parent population for which the program was implemented and from the parent

population for which the program was not implemented, and these segments are then assigned to the implementation group and the comparison group. This analysis technique measures the outturns of the outcome performance indicator for the implementation group and the comparison group over a given period (or at a particular point in time) after implementation of the program, and defines the impact of the program as the difference in the outturns between the two groups.

b. Technical limitations

The statistically equated model cannot be utilized for programs carried out uniformly nationwide because a group for which the program has not been implemented is needed for comparison purposes. The influence of external factors also cannot be completely eliminated because the implementation group and the comparison group are not completely homogeneous.

(8) Matching model

a. Techniques

To establish an implementation group and a comparison group under the matching model, an implementation group for whom the program is to be implemented and a comparison group with similar conditions as this implementation group in term of matching indicator for whom the program is not to be implemented are specified prior to the implementation of the program. This analysis technique measures the outturns of the outcome performance indicator for the implementation group and the comparison group over a given period (or at a particular point in time) after implementation of the program and defines the impact of the program as the difference in the outturns of the two groups. It is necessary in this analysis technique to confirm prior to the implementation of the program that the values for both the outcome performance indicator and the matching indicator are approximately the same for the implementation group and the comparison group.

b. Technical limitations

The matching model cannot be utilized for the program carried out uniformly nationwide because a group for which the program has not been implemented is needed for comparison purposes. The influence of external factors also cannot be completely eliminated because the implementation group and the comparison group are not completely homogeneous. In addition, the matching model cannot be used for the program that has already been implemented because the implementation group and the comparison group must be determined prior to implementation of the program.

(9) Regression-discontinuity model

a. Techniques

To establish an implementation group and a comparison group under the regression-discontinuity model, the parent population for which the program may be implemented is divided prior to implementation of the programs into a large group and a small group using a given standard value (cutoff point) based on the outturns of the outcome performance indicator. Next, the program is implemented for the small group (implementation group) and is not implemented for the large group (comparison group) – the reverse may also occur depending on the nature of the outcome performance indicator – with improvement in the outcome for the group smaller than the cutoff point regarded as necessary and improvement in the outcome for the group larger than the cutoff point deemed not necessary. This analysis technique measures the outturns of the outcome performance indicator for the implementation group and the comparison group at a particular point in time after implementation of the program, and defines the impact of the program as the difference in the regression lines of the outturns of the two groups at the cutoff point. It is necessary in this analysis technique to confirm prior to the implementation of the program that the regression line of the outturns of the outcome performance indicator for the

implementation group and the comparison group are continuous at the cutoff point.

b. Technical limitations

The regression-discontinuity model cannot be utilized for the program to be carried out uniformly nationwide because a group for which the program have not been implemented is needed for comparison purposes. The influence of external factors also cannot be completely eliminated because the implementation group and the comparison group are not completely homogeneous. In addition, the regression-discontinuity model cannot be used for the program that has already been implemented because the implementation group and the comparison group must be determined prior to implementation of the program.

(10) Randomized experimental model

a. Techniques

To establish an implementation group and a comparison group under the randomized experimental model, sample groups are randomly extracted prior to implementation of the program from the parent population for which the program could be implemented, and these sample groups are then randomly assigned to an implementation group for which the program is implemented or to a comparison group for which the program is not implemented. Because assignment to the implementation group and the comparison group is carried out randomly, the implementation group and the comparison group are for the most part homogeneous, except for the fact that the program will be implemented for the former and not for the latter. This analysis technique measures the outturns of the outcome performance indicator for the implementation group and the comparison group over a given period (or at a particular point in time) after implementation of the program and defines the impact of the program as difference in the outturns of the two groups. This analysis technique is regarded as one of the most reliable because the differences arising between the two groups after implementation of the program can all be attributed to the programs themselves.

b. Technical limitations

The randomized experimental model cannot be utilized for the program to be carried out uniformly nationwide because a group for which the program has not been implemented is needed for comparison purposes. This model can also not be adopted for the program already implemented because the implementation group and the comparison group must be established prior to implementation of the program. Furthermore, the reliability of the analysis results may decline as a result of a bias in the measurement results if the number of samples is reduced because agreement cannot be obtained from potential beneficiaries to participate in the implementation group or if participants in the implementation group drop out while the program is being implemented.

6. Analysis techniques for cost-efficiency analysis

Cost-efficiency analysis measures the cost-efficiency of the program to verify whether the program implemented by ministries/agencies had improvement effects on people's lives and the society/economy greater than the resources input. Key to doing so is the conversion of the improvement effects of the program into monetary values. The following analysis techniques are used to measure the cost-efficiency of the program.

(1) Cost-effectiveness analysis

a. Techniques

Cost-effectiveness analysis is an analysis technique that, once the social effects and social costs

stemming from implementation of the program have been identified, calculates these social effects and social costs without necessarily converting them all into monetary values and then compares the effects and costs indicated in a variety of units (monetary amounts, numbers of persons, numbers of cases, time periods, etc.). This analysis technique features methods for converting just the costs into monetary value and calculating the costs per unit of effect and the effect per unit of cost and methods that use numerical indicators, some of which are not based on monetary values, for both effects and costs. There are also methods for combining multiple effect and cost items into an appropriately weighted indicator. All else being equal, the best policy means from a cost-efficiency perspective is that which minimizes the costs per unit of effect and maximizes the effect per unit of cost.

b. Technical limitations

Cost-effectiveness analysis cannot verify simply by the analysis results whether the program in question had an improvement effect on people's lives and the society/economy greater than the resources input because the effects and costs are not necessarily converted into monetary values. When multiple policy means are employed to achieve the program, only relative cost and effect comparisons can be made between these policy means.

(2) Cost-benefit analysis

a. Techniques

Cost-benefit analysis is an analysis technique that, once the social benefits and social costs stemming from implementation of the programs have been identified, converts all benefit and cost items into monetary values for comparison. When benefits and costs arise continuously over a given period, comparisons are made after adjusting all benefits and costs to present values. Among methods of comparison are those that utilize the ratio of social benefits to social costs and those that utilize net social benefits, derived by subtracting social costs from social benefits. The program is seen as being socially rational from a cost-efficiency perspective when their social benefits exceed their social costs.

b. Technical limitations

Cost-benefit analysis requires the conversion of social benefits and social costs arising from implementation of the program into monetary values, but the difficulty of converting some of these social benefits and social costs into monetary values means that not all benefit and cost items can be calculated. Social benefits tend to be especially difficult to convert into monetary values due to the peculiar nature of government services, and variations in measurement preconditions, measurement items, and measurement techniques lead to greatly differing analysis results.

III. Implementation of comprehensive evaluations

1. Comprehensive evaluation efforts

The central-government reform carried out in Japan in January 2001 prompted the introduction of a policy evaluation system in the ministries and agencies, and the Government Policy Evaluation Act was enacted in June 2001 to heighten the effectiveness of this system. The Cabinet in December 2001 approved the Basic Guidelines on Policy Evaluation (hereinafter, "Basic Guidelines"), which allowed ministries/agencies to selectively use project evaluation, performance measurement, and/or comprehensive evaluation as an evaluation method. The Government Policy Evaluation Act went into force in April 2002, requiring ministries/agencies conducting policy evaluations to prepare evaluation reports listing the policies targeted for evaluation as well as the evaluation methods and policy effect analysis techniques used. Based on these evaluation reports[4], the present efforts and issues in comprehensive evaluations undertaken by ministries/agencies in FY2002 and FY2003 can be summarized as follows (see Table 1).

Table 1 Number of comprehensive evaluations implemented

(Unit: number of evaluations)

Ministry/agency	Basic Plan	FY2002	FY2003	Total
Cabinet Office	○	1	−	1
Imperial Household Agency	−	−	−	−
Japan Fair Trade Commission	○	1	1	2
National Public Safety Commission/National Police Agency	○	−	−	−
Defense Agency	○	16	11	27
Financial Services Agency	○	−	−	−
Ministry of Internal Affairs and Communications	○	−	1	1
Environmental Dispute Coordination Commission	−	−	−	−
Ministry of Justice	○	1	−	1
Ministry of Foreign Affairs	○	120	126	246
Ministry of Finance	○	1	−	1
Ministry of Education, Culture, Sports, Science and Technology	○	2	−	2
Ministry of Health, Labour and Welfare	○	1	3	4
Ministry of Agriculture, Forestry and Fisheries	○	1	−	1
Ministry of Economy, Trade and Industry	−	−	−	−
Ministry of Land, Infrastructure and Transport	○	11	8	19
Ministry of the Environment	−	−	−	−
Total	13 ministries/ agencies	155	150	305

Note) A ○ mark in the "Basic Plan" column indicates a ministry/agency that has selected comprehensive evaluations as one of evaluation methods in its Basic Plan.

4) The evaluation reports are publicly released on the websites of the individual ministries/agencies. From III.1. onward, the author uses data from the evaluation reports obtained from the websites of ministries/agencies.

(1) Present situation

a) Of the 17 ministries/agencies covered by the Government Policy Evaluation Act, 13 ministries/agencies[5] positioned comprehensive evaluation as one of evaluation methods in their respective Basic Plans.[6]

b) There were 10 ministries/agencies that actually conducted comprehensive evaluations in FY2002 and 6 in FY2003. Of these, 11 ministries/agencies carried out comprehensive evaluations in either FY2002 or FY2003, and 5 ministries/agencies undertook comprehensive evaluations during both of these fiscal years.

c) A total of 155 comprehensive evaluations were carried out in FY2002 and 150 in FY2003, with the total for both fiscal years being 305.

d) There are considerable differences in the number of comprehensive evaluations implemented by individual ministries/agencies; of the total of 305 such evaluations carried out in FY2002 and FY2003, the Ministry of Foreign Affairs conducted 246 (80.7%), the Defense Agency 27 (8.9%), the Ministry of Land, Infrastructure and Transport 19 (6.2%), and 8 other ministries/agencies four or fewer each.

e) In some instances ministries/agencies conduct a comprehensive evaluation by themselves while in others they contact out a part or whole of it to outside think tanks and organizations; 8 comprehensive evaluations were outsourced in FY2002 and 5 in FY2003 for a total of 13 during these two fiscal years.

(2) Issues

a. Need for undertaking comprehensive evaluations

At the majority of ministries/agencies, performance measurement is positioned as the core evaluation method of the three evaluation methods stipulated in the Basic Guidelines. Performance measurement is an evaluation method[7] that sets out in advance objectives to be achieved with a focus on policy effects and that measures performance vis-à-vis these objectives to evaluate the degree of success. Even if the evaluation concludes that the objectives have not been achieved, the performance measurement cannot determine if this is attributable to a lack of consistency in the logic model, a failure to implement the policy as planned, or the absence of effective policy means. On the other hand, even if the evaluation concludes that the objectives have been achieved, the performance measurement cannot establish whether the improvement effects on people's lives and the society/economy are greater than the resources input. As such analysis cannot be done without comprehensive evaluations, those ministries/agencies that carry out few or no comprehensive evaluations must actively undertake comprehensive evaluations.

b. Need for selecting limited number of topics

In some instances ministries/agencies appear to have attempted to cover an exhaustive range of policies under their jurisdiction in their comprehensive evaluations without selecting limited number of topics. The majorities of these evaluations simply describe the status of government activities with respect to the policies being evaluated, and offer almost no in-depth analysis. Comprehensive evaluations by their nature provide in-depth analysis from a variety of perspectives on the manifestation of policy effects in specific areas, ascertain problems with policies, and analyze their causes, without seeking to cover all of the policies implemented by ministries/agencies every fiscal year. In-depth analysis for comprehensive evaluations employing

5) The evaluation method termed program evaluation at the Ministry of Land, Infrastructure and Transport corresponds to comprehensive evaluation.

6) In these Basic Plans, ministries/agencies set out their fundamental provisions on policy evaluation, including evaluation objectives, evaluation methods, and evaluation standpoints, based on Article 6 of the Government Policy Evaluation Act and in line with the Basic Guidelines.

7) Basic Guidelines (appendix) [performance measurement method].

such evaluation techniques as theory evaluation, impact evaluation and cost-efficiency analysis is very costly and time-consuming. Consequently, ministries/agencies that have heretofore addressed an exhaustive range of their policies in comprehensive evaluations every year should instead take up specific topics on a priority basis and employ theory evaluation, impact evaluation, cost-efficiency analysis or other evaluation techniques to conduct in-depth analysis.

c. Need for focusing on policies

Some ministries/agencies focus their comprehensive evaluations not on policies themselves but rather on approaches to organizational operation. More specifically, they evaluate the reappointment system and other personnel matters, benefit packages such as the establishment of government residences, and confidentiality systems and other service disciplines. However, the Government Policy Evaluation Act stipulates that evaluations will examine policies,[8] and approaches to organizational operation themselves are not included within the scope of evaluation. Approaches to organizational operation are in the end reflected in the manifestation of policy effects as they constitute part of the infrastructure with which individual policies are implemented. Agencies/ministries that have thus far extensively examined their approaches to organizational operation in their annual comprehensive evaluations therefore need to address policies instead.

2. Evaluation techniques for comprehensive evaluations

Some of the evaluation techniques used in comprehensive evaluations by individual ministries/agencies are not strictly the same as those for program evaluation, but they can serve as adequate substitutes for the evaluation techniques in program evaluation depending on the process within the logic model that is the focus of attention. Matching the evaluation techniques utilized for the 305 comprehensive evaluations carried out in FY2002 and FY2003 by individual ministries/agencies to the evaluation techniques for program evaluation introduced in II.2. above gives the following table (see Table 2).

Table 2 Breakdown of evaluation techniques for comprehensive evaluation

(Unit: number of evaluations)

Evaluation technique	FY2002	FY2003	Total
Theory evaluation	18	29	47 (15.4)
Process evaluation	153	131	284 (93.1)
Impact evaluation	34	28	62 (20.3)
Cost-efficiency analysis	6	3	9 (3.0)
Total	211	191	402

Note 1) As multiple evaluation techniques have been used for each comprehensive evaluation in some cases, the row totals may not match those in Table 1.

Note 2) Figures in parentheses show the percentage of the total of 305 comprehensive evaluations indicated in Table 1.
As multiple evaluation techniques have been used for each comprehensive evaluation in some cases, the total does not add up to 100%.

(1) Present situation

a) Theory evaluation is an evaluation technique that focuses on the consistency of the logic model, and it was employed in 47 (15.4%) of the comprehensive evalutions. However, the majority of

8) Article 2.2 of the Government Policy Evaluation Act

these 47 evaluations did nothing more than illustrate the connections between policy objectives and policy means by flow charts and diagrams; almost none specifically examined the "input → activities → output → (external factors) → outcomes" logic model introduced in II.3 above and verified that the logic model had been designed consistent with the chain relation between cause and result.

b) Process evaluation is an evaluation technique that focuses on compliance with the logic model, and it was adopted in 284 (93.1%) of the comprehensive evalutions. However, the majority of these 284 evaluations did no more than describe the results of input, the results of government activities and the production results of output; almost none verified whether or not output was being provided according to plan as described in II.4.

c) Impact evaluation is an evaluation technique that focuses on the effectiveness of policy means, and it was used in 62 (20.3%) of the comprehensive evalutions. One or more of the analysis techniques introduced in II.5 were employed in these 62 impact evaluations.

d) Cost-efficiency analysis is an evaluation method that focuses on the cost-efficiency of policies, and it was utilized in 9 (3.0%) of the comprehensive evalutions. One or more of the analysis techniques introduced in II.6 were employed in these 9 cost-efficiency analysis.

e) Employing theory evaluation, process evaluation, impact evaluation and cost-efficiency analysis simultaneously in comprehensive evaluation would be ideal, and these four evaluation techniques were simultaneously used in 6 (2.0%) of the comprehensive evaluations.

(2) Issues

a. Need for enhancing and adopting theory evaluation

As the first stage of program evaluation, theory evaluation verifies the consistency of a logic model and in the process generates information essential to other evaluation techniques. If the logic model is not consistent, the improvement effects sought will not be achieved even if the policy is implemented as planned. A comparison is made in process evaluation between actual performance and the plan, and information regarding the plan can be ascertained through theory evaluation. An outcome performance indicator is used in impact evaluation to measure improvement effects, but information on this outcome performance indicator can be determined via theory evaluation. Consequently, ministries/agencies need to employ theory evaluation at a higher rate in conducting comprehensive evaluations and to pursue proper theory evaluation in order both to verify the consistency of the logic model and to ascertain information essential for other evaluation techniques.

b. Need for enhancing process evaluation

Process evaluation has been utilized in the vast majority of comprehensive evaluations, but the majority of these evaluations have only described the results of input, the results of government activities, and the production results of output. Almost none of these evaluations compared actual performance with plans in regard to the quantity/quality and timing of input and to the government activities of the implementing entity in order to verify whether or not the output was being delivered as planned, as discussed in II.4. The majority of process evaluations by individual ministries/agencies thus cannot rightly be called evaluations, as they do not constitute genuine process evaluations. Ministries/agencies need to incorporate proper process evaluation into their comprehensive evaluations to verify compliance with the logic model.

c. Need for employing impact evaluation

Impact evaluation is a technique used to evaluate whether a policy has had an improvement effect on people's lives and the society/economy. Even if a policy is implemented as planned, it will not have an improvement effect on people's lives and the society/economy unless effective policy means are utilized. While process evaluation does allow one to verify that a policy has been implemented as planned, it cannot alone ensure the manifestation of policy effects. Ministries/

agencies therefore need to make greater use of impact evaluation in their comprehensive evaluations to verify the effectiveness of policy means.

d. Need for using cost-efficiency analysis

Cost-efficiency analysis is a technique employed to evaluate whether a policy has had improvement effects on people's lives and the society/economy greater than the resources input. A policy cannot be considered socially rational from a cost-efficiency perspective if it does not have improvement effects on people's lives and the society/economy greater than the resources input. When there are multiple policy means for each policy objective, more resources must be input into policy means with higher cost-efficiency in order to enhance policy effects within the scope of limited resources. While impact evaluation can verify that a policy has had effects, it cannot alone ensure the continuity of the policy. Ministries/agencies therefore need to make greater use of cost-efficiency analysis in their comprehensive evaluations to verify the cost-efficiency of policies.

3. Analysis techniques for impact evaluation

The analysis techniques adopted by individual ministries/agencies for impact evaluation in comprehensive evaluations are the same as the analysis techniques used for impact evaluation in program evaluations. Matching the analysis techniques utilized for the 62 impact evaluations conducted by ministries/agencies in FY2002 and FY2003 with those analysis techniques introduced in II.5 produces the following breakdown (see Table 3).

Table 3 Breakdown of analysis techniques for impact evaluation

(Unit: number of evaluations)

Analysis technique		FY2002	FY2003	Total
Beneficiaries judgment		16	13	29 (46.8)
Simple before-after comparison model		11	13	24 (38.7)
Panel study		–	2	2 (3.2)
Interrupted time series model		13	11	24 (38.7)
Cross-section model		–	1	1 (1.6)
Semi-experimental model	Generic control model	4	2	6 (9.7)
	Statistically equated model	1	4	5 (8.1)
	Matching model	–	–	–
	Regression-discontinuity model	–	–	–
Randomized experimental model		–	–	–
Total		45	46	91

Note 1) As multiple analysis techniques have been used for each impact evaluation in some cases, the row totals may not match those in Table 2.

Note 2) Figures in parentheses show the percentage of the total of the 62 impact evaluations indicated in Table 2. As multiple analysis techniques have been used for each impact evaluation in some cases, the total does not add up to 100%.

(1) Present situation

a) Beneficiaries judgment was used in 29 (46.8%) of the impact evaluations. Beneficiaries judgment is regarded as the least reliable in terms of analysis results, but it is nevertheless the most commonly used technique for impact evaluation.

b) The simple before-after comparison model was utilized in 24 (38.7%) of the impact evaluations.

c) Panel studies were adopted in 2 (3.2%) of the impact evaluations.

d) The interrupted time series model was employed in 24 (38.7%) of the impact evaluations. Visual analysis by table and graph was performed in 21 cases, and statistical analysis by regression analysis in 3 cases.

e) The cross-section model was used in 1 (1.6%) of the impact evaluations. This entailed visual analysis by graph.

f) The generic control model was utilized in 6 (9.7%) of the impact evaluations. The generic control model compares with the nationwide average or the average across all industries.

g) The statistically equated model was employed in 5 (8.1%) of the impact evaluations.

h) In general, analysis techniques not requiring a comparison group were adopted quite often. Although analysis techniques featuring comparison groups were used, no use was made of analysis techniques that involved establishing comparison groups prior to policy implementation.

(2) Issues

In impact evaluation, the differences in the outcome performance indicator between the implementation group and the comparison group and within the implementation group before and after policy implementation are regarded as the effectiveness of policy means. Because the impact of external factors is included within these differences, it is important to adopt analysis techniques that remove the impact of these external factors in order to improve the reliability of the analysis results.

a. Need for establishing a comparison group

The first stage in eliminating the impact of external factors is to use as far as possible analysis techniques that require the use of a comparison group, as the implementation group cannot maintain the homogeneity of elements other than the implementation of policies (or lack thereof) over an extended period. Of the analysis techniques used by individual ministries/agencies in impact evaluation, those requiring a comparison group were used in only a fraction of the impact evaluations, even when the policies being evaluated were not implemented uniformly nationwide, i.e., even when it was possible to employ analysis techniques that utilize comparison groups. Therefore, ministries/agencies need to utilize analysis techniques that use comparison groups for policies not implemented uniformly nationwide in order to improve the reliability of analysis results.

b. Need for establishing the comparison group prior to policy implementation

The second stage in removing the impact of external factors is to use as far as possible analysis techniques that establish comparison groups prior to policy implementation. In many instances, the homogeneity of the comparison group with the implementation group before policy implementation cannot be confirmed once policies have been implemented. The entities conducting impact evaluations for individual ministries/agencies, excepting cases where such tasks have been outsourced to outside think tanks, have for the most part been policy implementation departments and bureaus. It is thus possible to adopt analysis techniques that establish comparison groups prior to policy implementation. Ministries/agencies need to employ analysis techniques requiring comparison groups prior to policy implementation whenever impact evaluations are conducted by policy implementation departments and bureaus in order to improve the reliability of analysis results.

4. Analysis techniques for cost-efficiency analysis

The analysis techniques used by individual ministries/agencies for cost-efficiency analysis in comprehensive evaluations are the same as the analysis techniques used for cost-efficiency analysis in program evaluations. Matching the analysis techniques adopted for the 9 cost-efficiency analysis carried out by individual ministries/agencies in FY2002 and FY2003 to the analysis techniques introduced in II.6 gives the following breakdown (see Table 4).

Table 4 Breakdown of analysis techniques for cost-efficiency analysis

(Unit: number of evaluations)

Analysis technique	FY2002	FY2003	Total
Cost-effectiveness analysis	1	1	2 (22.2)
Cost-benefit analysis	6	2	8 (88.9)
Total	7	3	10

Note 1) As multiple analysis techniques have been used for each cost-efficiency analysis in some cases, the row totals may not match those in Table 2.

Note 2) Figures in parentheses show the percentage of the total of the 9 cost-efficiency analysis indicated in Table 2. As multiple analysis techniques have been used for each cost-efficiency analysis in some cases, the total does not add up to 100%.

(1) Present situation

a) Cost-effectiveness analysis was used in 2 (22.2%) of the cost-efficiency analysis.

b) Cost-benefit analysis was used in 8 (88.9%) of the cost-efficiency analysis.

(2) Issues

One reason that cost-efficiency analysis was the least commonly utilized of the four evaluation techniques is that cost-benefit analysis cannot be conducted due to the difficulty of converting policies' social benefits into monetary values. However, cost-effectiveness analysis of policy means can be performed if it can be verified by impact evaluation that multiple policy means produce the same improvement effect, even if the social benefits of a policy cannot be converted into monetary values. This makes it possible to allocate greater resources to policy means having greater cost-efficiency. In cases where impact evaluation can verify that multiple policy means have had the same improvement effect, ministries/agencies need to conduct cost-effectiveness analysis through cost-efficiency analysis in order to ascertain the more highly cost-efficient policy means.

IV. Conclusion

The project evaluation method, the performance measurement method, and the comprehensive evaluation method are to be used in Japan for policy evaluation. Because performance measurement is one of the simpler among these, it is considered a core evaluation method for policy evaluation by individual ministries/agencies because it enables ministries/agencies to exhaustively examine the major policies under their jurisdiction. Even if performance measurement can provide information on the degree of progress made toward achieving objectives, however, it cannot provide information on the causes when objectives are not achieved nor can it provide information on the cost-efficiency of policy means. This information can be uncovered by theory evaluation, process evaluation, impact

evaluation and cost-efficiency analysis in the context of comprehensive evaluations.

Comprehensive evaluation can accordingly be seen as an important evaluation method complementing performance measurement, but process evaluation – and not even genuine process evaluation in most cases – continues to be the key approach taken by individual ministries/agencies. Hence, ministries/agencies must enhance their comprehensive evaluations referring to the techniques for program evaluation.

(Reference)

Rossi, Peter H., Lipsey, Mark W. and Freeman, Howard E. (2004) *Evaluation: A Systematic Approach*, 7th edition, Thousand Oaks: Sage Publications

RYU Yoshiaki and SASAKI Ryo (2004), *Theory and Techniques of "Policy Evaluation"* (enlarged and revised edition), Taga Shuppan

SASAKI Ryo (2003), 'Practical Examples of Program Evaluation,' *NIRA*, 2003, Vol.16 No.5

SASAKI Ryo (2003), *Policy Evaluation Training Book*, Taga Shuppan

TANABE Tomoko (2002), 'Policy Evaluation Techniques – Based on Evaluation Theory and Practice in the US,' *Administrative Management Research Quarterly*, No. 97

TANABE Tomoko (2003), 'Overview of Program Evaluation Techniques,' *NIRA*, 2003, Vol.16 No.5

United States General Accounting Office (1991), *Designing Evaluations* (GAO/PEMD-10.1.4)

United States General Accounting Office (1998), *Performance Measurement and Evaluation - Definitions and Relationships* (GAO/GGD-98-26)

Weiss, Carol H. (1998), *Evaluation: Methods for Studying Programs and Policies*, 2nd edition, Upper Saddle River: Prentice-Hall

Wholey, Joseph S., Hatry, Harry P. and Newcomer, Kathryn E. eds. (1994), *Handbook of Practical Program Evaluation*, San Francisco: Jossey-Bass

YAMADA Harunori (2000), *Policy Evaluation Techniques*, Nippon Hyoronsha Co., Ltd.