Part 4

Secondary Evaluation by the Advisory Committee on Evaluation

Secondary Evaluation by the Advisory Committee on Evaluation

Masafumi Kuroki
Vice-President
Chairperson of JICA Evaluation Study Committee

JICA established the Advisory Committee on Evaluation in fiscal 2002 and since then has committed itself to enhancing the evaluation system and improving projects using evaluation results, while receiving advice from the Committee. As part of that effort and in order to increase transparency and objectively in evaluation results, the Advisory Committee on Evaluation has evaluated terminal evaluations conducted by JICA (secondary evaluation), the results of which have been published in the Annual Evaluation Report since fiscal 2003. This fiscal year as well, with the help of the Japan Evaluation Society, the Advisory Committee on Evaluation set up a working group consisting of third-party experts in evaluation to conduct secondary evaluations. The results of the secondary evaluations are presented on the following pages.

This year's secondary evaluation first examined the quality of terminal evaluations (primary evaluation) conducted by JICA in fiscal 2005 and 2006. Also, based on the information contained in the terminal evaluation reports, individual projects were evaluated by the working group. In addition, a new attempt was made to rate each project based on its evaluation results.

Regarding the quality of primary evaluation, its results show that all nine evaluation criteria received more than three points on a scale of five. Also, improvements in quality were observed when comparing the primary evaluation results of fiscal 2004, 2005, and 2006 with those of 2003. Nonetheless, the areas that JICA must improve were pointed out, such as more concretely expressing the lessons learned from the promoting and impeding factors, and having overseas offices make more efforts to increase evaluation quality.

With respect to the quality of projects, the average scores for all five evaluation criteria reached more than three on a scale of five. The projects in fiscal 2004, 2005, and 2006 were graded higher than those in fiscal 2003. Moreover, the results of ratings based on the evaluation of projects in each fiscal year clarified that projects with a rating of B or higher have increased annually, while those with a rating of D or lower have decreased. As the points necessary to increase project quality, clarifying the causal relationship between the project purpose and overall goal, and setting appropriate indicators and target values when planning a project were recommended.

Considering the secondary evaluation results examined from the viewpoint of a third party, JICA will take further steps to implement more effective and efficient projects, and improve project evaluation.

Last but not least, I would like to express my sincere gratitude to every member of the Advisory Committee on Evaluation and its working group for offering valuable comments and recommendations. The members carefully examined 50 terminal evaluation reports (67 reports when including the previous years) from various aspects and exercised their ingenuity in conducting secondary evaluation.

Improving JICA's Evaluations and Projects (Conclusion)

Hiromitsu Muta
Chairperson of the Secondary Evaluation Working Group
Chairperson of the Advisory Committee on Evaluation

1. Overall Assessment

The Advisory Committee on Evaluation conducted secondary evaluations on terminal evaluations as it did in fiscal 2006, and also carried out an overall evaluation (rating) of all projects in fiscal 2007. The following conclusions are drawn from the analysis results.

(1) Maintaining the Quality of Evaluation

Secondary evaluators in fiscal 2007 differ from their predecessors. However, the secondary evaluation in fiscal 2007 reconfirmed that the quality of terminal evaluations improved significantly from fiscal 2003 through 2004 and generally maintained a high level even after fiscal 2004.

(2) Improving the Results of Project Evaluation

A chronological comparison was made of the evaluation results of secondary evaluators concerning the terminal evaluations in fiscal 2003 to 2006. Changes in average scores indicate that statistically, the projects in fiscal 2004, 2005, and 2006 are significantly higher than those in fiscal 2003. The projects in fiscal 2004, 2005, and 2006 do not show statistically significant differences, but do indicate an upward trend. We considered the sum of the scores weighted by the DAC's five evaluation criteria as the overall evaluation for five levels of rating:

A (excellent project)
B (good project)
C (fair project)
D (partially weak project)
E (weak project)

Then by looking at the distribution of ratings for projects, we found that as the years go by, projects rated B or higher increased, while projects rated D or lower decreased. Although the ratings were assigned by secondary evaluators based on the terminal evaluation reports, it is presumed that the projects have been evaluated high as years go by.

2. Major Evaluation Results

(1) Quality of Terminal Evaluation

In every evaluation criterion considered important regarding quality of evaluation, the scores averaged 3.2 or higher, exceeding the "medium" level of 3.0, thus ensuring high quality in terminal evaluation. Of those criteria, the evaluations of "data collection" and "recommendations" are relatively high, while those of "lessons learned" and "reporting" are relatively low.

A detailed look at the evaluation criteria by viewpoint reveals that "evaluability" (preconditions for appropriate evaluation) is evaluated at or above the "medium" level in every viewpoint. Among those viewpoints, "evaluability of overall goal" is evaluated with the lowest score. This low score is considered due to vague overall goals and indicators, or because several projects have unclear causal relationships between their project purposes and overall goals.

The "evaluation framework" is judged from the viewpoints of evaluation team composition and counterpart participation. The secondary evaluation in fiscal 2006 gave this evaluation criterion a relatively low score, but the present survey indicated that both viewpoints were evaluated at or above the "medium" level. In particular, "evaluation team composition" is rated higher than the viewpoints in other evaluation criteria. Since more reports now list the name and expertise of evaluators on the part of partner countries, it becomes easier for secondary evaluators to monitor quality of the evaluation team.

All the viewpoints in "Data collection" show no variances in evaluation and achieve higher scores than other evaluation criteria.

As for "assessment of performance," great variances are shown in the evaluation of viewpoints. "Outputs" is the most highly evaluated among all viewpoints, including other evaluation criteria. However, the evaluation of "overall goal" failed to reach even the "medium" level, thus receiving the lowest evaluation of all viewpoints. This low evaluation is due to insufficient assessment of performance and verification caused by vague overall goal and indicators.

"Analysis method" ensures an evaluation at or above the "medium" level in every viewpoint. Among those viewpoints, “analysis of promoting and impeding factors” is given a relatively low evaluation.

Each viewpoint in "evaluation" is at or above the "medium" level, but there are variances in the evaluation of viewpoints. Among all viewpoints, "relevance" is given the highest evaluation, even relatively higher than the viewpoints in other evaluation criteria. On the other hand, "efficiency" is given the lowest evaluation, even lower than the viewpoints in other evaluation criteria. Many reports evaluate efficiency from the viewpoints of the implementation process, such as the dispatch schedules of experts and the utilization of equipment. Therefore, few reports give an evaluation in comparison with similar projects or from the viewpoint of cost-effectiveness. The low evaluation is also attributed to insufficient information about actual expenses.

"Recommendations" ensures a high evaluation in every viewpoint, with little variances in evaluation.

"Lessons learned" ensures an evaluation at or above the "medium" level in every viewpoint. In contrast, "sufficiency of recommendations" is given a relatively low evaluation since lessons learned in some projects are too general and insufficient in depth.

"Reporting" obtains an evaluation at or above the "medium" level in every viewpoint. Evaluation of "presentation/legibility and clarity" is given a relatively high evaluation, with reports being written clearly. However, there are not many persuasive reports which use diagrams effectively and present primary data.

(2) Year-to-Year Changes in the Quality of Terminal Evaluation

A closer look at year-to-year changes in the quality of terminal evaluation revealed that, in analyzing secondary evaluations conducted in fiscal 2005 and 2006, the terminal evaluations in fiscal 2004 and 2005 are higher in quality than the terminal evaluation in fiscal 2003. The present analysis indicated that the quality of terminal evaluation in fiscal 2006 is also higher than that of fiscal 2003. Thus, the quality of the terminal evaluation in fiscal 2004 and later is higher than that in fiscal 2003, and it is still maintained. On the other hand, year-to-year changes among fiscal 2004, 2005, and 2006 are unclear.

(3) Differences in Quality between Evaluations by JICA Headquarters and Overseas Offices

It was found that the quality of terminal evaluation conducted by overseas offices is lower than that of headquarters, although as few as eight evaluations were conducted by overseas offices. The terminal evaluation by overseas offices had a weakness in logical comprehension and somewhat insufficient analysis, which apparently led to the low quality of evaluation. Among the projects implemented by overseas offices, some problems on "evaluability" were found in projects conducted in collaboration with other donors and projects with peculiar external factors. This fact presumably contributed to the low quality of terminal evaluation by overseas offices.

(4) Project Evaluation by Secondary Evaluators Based on Terminal Evaluation Reports

Projects targeted for secondary evaluation received the "medium" level or above in every evaluation criterion in the DAC's five evaluation criteria. Of those, "relevance" is most highly evaluated, while "efficiency" is the lowest.

In "relevance," which is generally given a high evaluation, the viewpoints of "validity of project implementation" and "necessity of the project" are highly evaluated. However, "appropriate approach," which questions whether the approach is appropriate to solve the development issues, is relatively low.

"Effectiveness" is highly evaluated in every viewpoint of "achievement level of project purpose" and "causal relationships between outputs and project purpose." It is given a higher evaluation than the viewpoints of other evaluation criteria as well.

"Efficiency" is highly evaluated in the viewpoint of "clear input cost." However, low evaluation is given to the viewpoints of "cost-effectiveness," which questions whether utmost efforts are made for cost-effectiveness, and "appropriate implementation process," which questions the appropriateness of timing and scale, and fail to reach the "medium" level.

"Impact" is evaluated at or above the "medium" level in every viewpoint. In addition, "unintended positive and negative impact" is given a higher evaluation than the viewpoints of other evaluation criteria.

"Sustainability" is evaluated at or above the "medium" level in every viewpoint. However, a relatively low evaluation is given to the "mechanism of securing sustainability," which measures whether the mechanism of securing sustainability is considered in a project.

(5) Differences in Evaluations between Projects by JICA Headquarters and Overseas Offices

Projects implemented by headquarters generally tend to be somewhat higher in project evaluations by secondary evaluators than projects implemented by overseas offices. Above all, the scores of "appropriate approach" concerning "relevance" and "causal relation between outputs and project purpose" concerning "effectiveness" are statistically higher in projects implemented by headquarters. Projects by headquarters clearly present their purposes and set up appropriate plans toward achieving those purposes when the projects are launched, thereby leading to the projects achieving high scores.

3. Toward Further Improvements

The following are specific items extracted from the results of analysis for the improvement of quality in both evaluation and the project itself.

(1) For Improvement of the Quality of Primary Evaluation

1) Evaluation team composition: Increase counterpart participation in the evaluation, and specify in the report to what extent the evaluators take part in both the projects and evaluation.

2) Formulation of appropriate PDM: In preparing a PDM, clarify the indicators and causal relationships from outputs to the project purposes and overall goals.

3) Data collection: If indicators to measure the achievement of project purpose cannot be obtained by the means specified in the PDM, make efforts to obtain the information by alternative means. To increase the objectivity of evaluation, widen the sources of data collection. To conduct analysis from the viewpoint of cost-effectiveness, collect data to enable a cost comparison with similar projects.

4) Objective analysis: To ensure objectivity, clearly specify the process leading to and the grounds for the conclusion.

5) Recommendations and lessons learned: Extract recommendations and lessons learned sufficiently from the inhibiting factors, and present them concretely, considering the utilization of recommendations as measures for project improvement and lessons learned in similar projects.

6) Reporting: Remember that the public will read the reports. Therefore, write the evaluation results clearly to maintain objectivity. Also use tables and figures to make reports easy to understand for readers.

7) Headquarters will give the overseas offices guidance on the evaluation methodology inclusive of the formulation of PDM. The overseas offices ensure that evaluations are implemented according to the JICA guidelines.

(2) For Improvement of the Quality of Projects

1) Appropriate approach for project implementation: In project implementation, relevance at the initial stage will affect the subsequent implementation processes and outcomes. So, carefully study external factors and the situation surrounding the project. Then, select appropriate means and methods.

2) Causal relationships between the project purpose and overall goal: The outcome defined as the overall goal emerges after the project purpose is achieved. To increase achievement of the overall goal, clarify the causal relationships from outputs to project purpose and overall goal.

3) Setting indicators and target values: In designing a PDM, it is important to set appropriate indicators corresponding to activities and project purpose, and set target values for objectivity in the results.

4) Guidance to overseas offices on how to implement a project: Headquarters takes measures to improve the quality of projects implemented by overseas offices, such as the provision of guidance to the overseas offices on the formulation of PDM and evaluation methodology. The overseas offices, in turn, ensure the preparation of an appropriate PDM in line with JICA guidelines.

Results of Secondary Evaluation in Fiscal 2007

Advisory Committee on Evaluation/
Secondary Evaluation Working Group

1. Objectives, Targets, and Methods of Secondary Evaluation

(1) Objectives of Secondary Evaluation

Who should evaluate ODA projects? There might be a number of potential evaluators. For evaluations made by stakeholders, a detailed evaluation can be expected in light of circumstances, since the evaluators have profound knowledge of the project and region, and fully understand the activities and various situations involved. Feedback is also more likely to fully function, leading to improvements in the project. Conversely, more lenient evaluations could result since evaluators may give too much allowance for circumstances, leading to problems in transparency and impartiality. Due partly to the nature of its operation, JICA manages a number of relatively small-scale projects, and for the terminal evaluation alone, there are around 50 projects every year. Therefore, in reality, JICA can only conduct internal evaluations; otherwise, it must seek the assistance of outside stakeholders, such as supporting committee members, to conduct the evaluation.

As a means of overcoming the expected disadvantages while taking advantage of internal evaluation, objectivity and impartiality can be accordingly achieved by thoroughly conducting a project evaluation in compliance with guidelines, and through secondary evaluation by external experts based on internal evaluation results. This secondary evaluation does not aim to reevaluate individual projects, but to grasp the general trend of the quality of terminal evaluations and suggest ways of improvement.

The Plan-Do-Check-Act (PDCA) cycle is an effective tool to continually improve projects. Evaluation corresponds to the Check part of this cycle. When applying the concept to improving evaluation, the PDCA cycle of evaluation will be referred to as the planning of evaluation, implementation of evaluation, evaluation of evaluation, and improvement of evaluation. In order to diminish the bias in evaluation, it is important to incorporate the views of external examiners. However, they need not actually evaluate every single project. At the very least, a certain level of transparency and objectivity can be secured by incorporating the views of external examiners into the Check part of the PDCA cycle.

Evaluation involves a series of processes for collecting information, conducting analysis, drawing recommendations/lessons, and compiling reports based on an evaluation framework. In order to ensure the reliability of primary evaluation, such as the terminal evaluation of projects as in previous years, and facilitate the disclosure of easy-to-understand evaluation results, the secondary evaluation in fiscal 2007 was conducted by focusing on the following questions:

a. Evaluation of the quality of primary evaluation
- Is the primary evaluation sufficiently qualified?
- Has the quality of primary evaluations improved annually?
- What tasks should be conducted to further upgrade the quality?

b. Evaluation of projects by secondary evaluators based on reports (i.e., primary evaluation)
- What is the result of secondary evaluation of the project?
- Have project evaluation results improved annually?
- What are the factors that influence project evaluation results?

(2) Evaluators

The question is raised about who conducts secondary evaluation. The principle of secondary evaluation refers to whether the evaluation results are convincing to stakeholders, rather than whether they are correct or not. There is no single answer to this question about how evaluation should be conducted, and answers vary depending on the evaluator's background and sense of value that affect the evaluation. When target values are set for projects, it is easier to agree as to whether project purposes have been achieved or not. It is still natural that there are gaps in opinions on the measures to be taken. There is no guarantee that the results of secondary evaluation conducted by secondary evaluators are the utmost

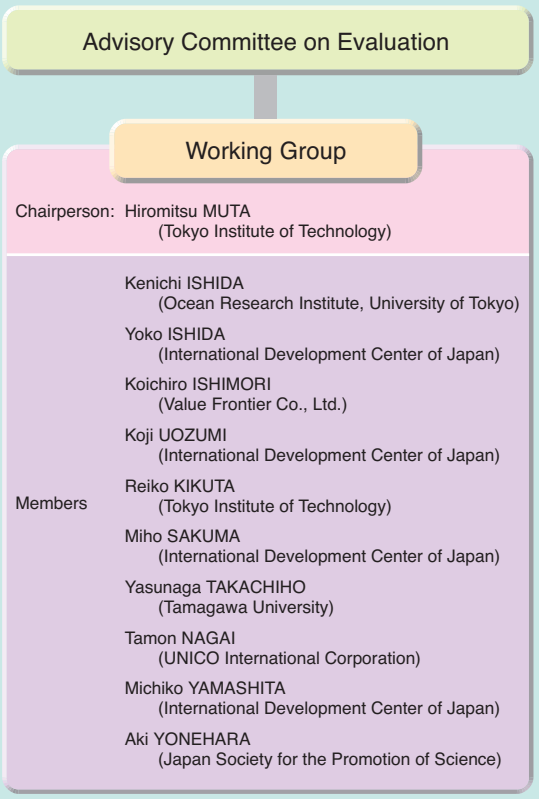
and foremost. It is likely that such results will differ when another evaluator conducts secondary evaluation. In such cases, it is safer and more practical to devise a framework to allow the opinions of several secondary evaluators with some level of ability, rather than evaluation by one outstanding evaluator.

JICA established the Advisory Committee on Evaluation to solicit opinions on the nature of JICA evaluations and results. However, due to the nature of the committee, the opinions there tend to be too general, making it difficult to conduct a detailed secondary evaluation of each evaluation result. Thus, it is practical to set up a working group to conduct secondary evaluation by taking time to scrutinize internal evaluation results and discuss outcomes at the meetings of higher-level committees.

In fiscal 2003, JICA launched the secondary evaluation of terminal evaluation by setting up its Working Group. Eight members of the Advisory Committee on Evaluation took charge of the secondary evaluation in fiscal 2003. In fiscal 2004, the Secondary Evaluation Working Group was formed under the Advisory Committee on Evaluation, consisting of six experts and eight JICA staff members (primarily evaluation chiefs of each department). The experts were selected based on recommendations by the Japan Evaluation Society to guarantee the objectivity of selection. Differences in evaluation tendencies between the experts and JICA staff were also observed, and it was concluded that no significant difference exists in such evaluation tendencies.

Since the methodology was sufficiently developed for practical application in fiscal 2004, the task of secondary evaluation was subcontracted to the Japan Evaluation Society in fiscal 2005 and 2006. In fiscal 2007, the task was also subcontracted to the Japan Evaluation Society, which formed a 10-member evaluation team. The members were recruited within the Society for transparency. Figure 4-1 illustrates the secondary evaluation system of fiscal 2007.

Figure 4-1 Secondary Evaluation System



(3) Evaluation Targets

The secondary evaluation in fiscal 2007 targeted 25 terminal evaluations conducted in fiscal 2005 and 25 in fiscal 2006. All 50 projects were the subjects of study. Moreover, for a year-to-year comparison, three evaluations in fiscal 2003, 11 in fiscal 2004, and three in fiscal 2005 among 65 terminal evaluations subject to the secondary evaluation conducted in fiscal 2006 were selected for secondary evaluation in fiscal 2007 (Appendix 1).

(4) Evaluation Design and Methods

If all study members can study all the evaluation reports, the mean scores for each evaluation criterion (evaluation viewpoint/criterion) will reflect the opinions of all evaluators, with the results being free of bias. This is because the results would be biased unless the opinions of a certain number of evaluators are averaged, since each evaluator has a different background and opinion. However, in reality, the actual workload placed on each evaluator is too enormous to bear alone. For example, it takes at least two to three hours for an evaluator to thoroughly read a report and fill in the scores and comments on an evaluation sheet. There are about 70 reports subject to the secondary evaluation for this year, including those taken over from the previous year. Even if possible, this poses an excessive burden for the working group.

Thus, in fiscal 2007, four different evaluators read each evaluation report. Specifically, one member of the working group read all 67 reports, one member read 34 reports, one read 33 reports, another read 20 reports, and the other six members read 19 reports. This scheme allows us to treat the judgment criteria of one key member as the norm for the entire group and adjust the judgment criteria of other evaluators. Although fairness is more likely with four evaluators rather than one, it is still unavoidable that the tendencies of specific evaluators could affect the results. Theoretically, the scores given by each evaluator can be divided into two parts: the true score for the evaluation target (free of the evaluator's personal evaluation bias) and the coefficient of evaluation tendency for each evaluator (strictness/leniency coefficient: error tendencies of individual evaluators). Accordingly, as in previous years, a method of statistical analysis (to analyze variance) was employed to differentiate both parts so that the evaluation tendencies of evaluators could be adjusted to obtain an unbiased estimate of evaluation scores that are free of personal evaluation tendencies.

A comparative study was conducted year-by-year by sam-

pling the series of project evaluations over the years. These projects that had been evaluated repeatedly can serve as the so-called "seam allowance" for equating. By using the seam, it is possible to link the secondary evaluation results of fiscal 2006 and 2007. True estimates of the evaluation scores were calculated for fiscal 2006 and 2007; however, the evaluation standard itself may be different. In order to see the distribution of evaluation scores, it is effective to match both the mean scores and variances over two years based on the seam by conversion. Specifically, all secondary evaluation results obtained before fiscal 2006 should be converted so that the mean scores and variances for each fiscal year correspond to each other. With proper sampling for the seam, such a simple conversion is sufficient for making a comparison. In this way, evaluation data obtained individually can be processed and analyzed as a large pooled sample by equating disconnected evaluation information in various ways using the seam allowance.

(5) Evaluation Sheet Structure and Analysis Methods

The secondary evaluation of terminal evaluation has two objectives: evaluating the quality of terminal evaluation, and checking the quality of a project using the terminal evaluation.

In a secondary evaluation, experts basically evaluate the evaluation results (reports) based on a set of evaluation viewpoints. Evaluation items listed on the evaluation sheet and in the criteria were based on the criteria for good evaluations in the Revised JICA Evaluation Guidelines (March 2004).

The evaluation sheet for fiscal 2007 is mostly the same as that for fiscal 2006. The five-point rating scale makes it possible for a year-to-year comparison of the results in fiscal 2005, 2006, and 2007. Table 4-1 lists the changes in evaluation viewpoints from fiscal 2004 to 2007.

Table 4-2 and Appendix 2 give the evaluation viewpoints employed in fiscal 2007. The following section describes analysis conducted based on these evaluation viewpoints. Evaluation was conducted based on the following five-point scale for rating both viewpoints and scoring:

5: Sufficient/high
4: Fairly sufficient/high
3: Average
2: Slightly insufficient/low
1: Insufficient/low

Also reported is that in fiscal 2007, the secondary evaluation contains an additional overall score that comprehensively evaluates the quality of the project apart from specific results of the DAC's five evaluation criteria scored by secondary evaluators from the report. Moreover, to comprehensively evaluate the project by secondary evaluators based on the report, we weighted each score of the DAC's five evaluation criteria and calculated the weighted overall score, which is a sum of each score. Adjustments are made so that the sum of weights given to each criterion becomes 1.0. Therefore, the minimum of weighted overall score is 1.0, while the maximum is 5.0. The equation used is as follows:

Weighted overall score
= (score for relevance x 0.13) + (score for effectiveness x 0.27) + (score for efficiency x 0.20) + (score for impact x 0.20) + (score for sustainability x 0.20)

Based on the result of questionnaire to members of the Advisory Committee on Evaluation, the weights of each criterion are given by the average scores among committee members.

Table 4-1 Comparison of Evaluation Viewpoints and Rating Scale between Fiscal 2004, 2005, 2006, and 2007

Fiscal Year	2004		2005		2006		2007	
Evaluation Criteria	Viewpoints	Rating	Viewpoints	Rating	Viewpoints	Rating	Viewpoints	Rating
Evaluability	4 (+)	**	4 (*)	*	6 (*)	*	6 (*)	*
Evaluation Framework	4 (+)	**	3 (*)	*	2 (*)	*	2 (*)	*
Data Collection	5 (+)	**	4 (*)	*	4 (*)	*	4 (*)	*
Assessment of Performance	4 (+)	**	4 (*)	*	6 (*)	*	6 (*)	*
Analysis Method	3 (+)	**	3 (*)	*	3 (*)	*	3 (*)	*
Evaluation	7 (+)	**	6 (*)	*	6 (*)	*	6 (*)	*
Recommendations	4 (+)	**	3 (*)	*	3 (*)	*	3 (*)	*
Lessons Learned	4 (+)		3 (*)		3 (*)	*	3 (*)	*
Reporting	4 (+)	**	3 (*)	*	3 (*)	*	3 (*)	*
General Criteria for Good Evaluation	4 (+)	**						
Project Evaluation: Relevance		**	3 (*)	*	3 (*)	*	3 (*)	*
Project Evaluation: Effectiveness		**	2 (*)	*	2 (*)	*	2 (*)	*
Project Evaluation: Efficiency		**	2 (*)	*	3 (*)	*	3 (*)	*
Project Evaluation: Impact		**	3 (*)	*	3 (*)	*	3 (*)	*
Project Evaluation: Sustainability		**	5 (*)	*	5 (*)	*	5 (*)	*
Project Evaluation: Overall Evaluation		**						*

Evaluation method +: 3-point scale, *: 5-point scale, **: 10-point scale

2. Quality of Terminal Evaluation Examined through Reports

(1) Overview of Evaluation Results

The secondary evaluation in fiscal 2006 targeted 45 terminal evaluations (28 evaluations in fiscal 2004 and 17 in fiscal 2005). This fiscal year, the quality of 50 terminal evaluations was assessed (25 in fiscal 2005 and 25 in fiscal 2006) after the last evaluation. Figure 4-2 shows the average scores for individual evaluation criteria. The average scores of nine criteria (i.e., “evaluability,” “evaluation framework,” “data collection,” “assessment of performance,” “analysis method,” “evaluation,” “recommendations,” “lessons learned,” “reporting”) are 3.0 or higher. There are some differences between each evaluation criterion in terms of evaluation level. Scores are relatively high for "data collection” in conducting the evaluation, “assessment of performance" for analysis, and "recommendations" for the future. However, the average scores are relatively low for "evaluability" to see whether an appropriate evaluation is possible, "analysis method" and "lessons learned" which are obtained from the projects, and "reporting."

As for the distribution of scores shown in Figure 4-3, the scores of many evaluation criteria are distributed between 2.5 and 4.0. Many projects scored 3.0 or higher for “evaluability,” “evaluation framework,” “data collection,” “assessment of performance,” “analysis method,” “evaluation,” and “recommendations.” For “lessons learned” and “reporting,” half the projects scored less than 3.0 and the other half 3.0 or higher, while most projects scored 2.5 to 2.99 for “lessons learned.” Some projects scored 4.5 or more for "reporting, however, scores generally vary with a concentration on the low side.

It can be concluded that the quality of terminal evaluations tends to be at a level higher than “medium” on the grading scale.

Figure 4-2 Score Results by Evaluation Criterion (Average)

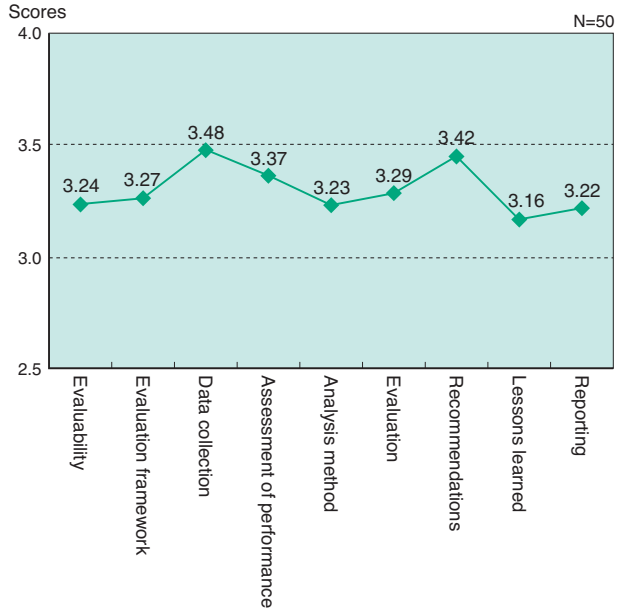


Table 4-2 Secondary Evaluation Criteria

I Criterion: The precondition for conducting appropriate evaluation was possible (Evaluability).
Viewpoints:
- Evaluability of Project Plan (Preliminary Study/PDM)
- Target Group
- Evaluability of Project Purpose
- Evaluability of Overall Goal
- Logic of Project Design
- Project Monitoring

II. Key Evaluation Criteria

1. Criterion: Evaluation Framework
Viewpoints:
- Time Frame of Evaluation Study
- Evaluation Team Composition—Impartiality and Specialty
- Level of Counterpart Participation

2. Criterion: Data Collection
Viewpoints:
- Evaluation Questions
- Appropriateness of Data Collection Methods and Data Sources
- Data/Information Sources
- Sufficiency of Data/Information Obtained

3. Analysis

3.1 Criterion: Assessment of Performance
Viewpoints:
- Measurement of Results (Outputs)
- Measurement of Results (Project Purpose)
- Measurement of Results (Overall Goal)
- Examination of Project Implementation Process
- Examination of Qualitative Causal Relations—Logic of Project Design
- Examination of Quantitative Causal Relations—Before and After

3.2 Criterion: Analysis Method
Viewpoints:
- Objective Analysis
- Holistic Analysis
- Analysis of Promoting and Impeding Factors

4. Criterion: DAC’s Five Evaluation Criteria
Viewpoints:
- Relevance
- Effectiveness
- Efficiency
- Impact
- Sustainability
- Conclusion

5. Recommendations/Lessons Learned

5.1 Criterion: Recommendations
Viewpoints:
- Sufficiency of Recommendations
- Relevance and Credibility of Recommendations
- Usability of Recommendations

5.2 Criterion: Lessons Learned
Viewpoints:
- Sufficiency of Lessons Learned
- Relevance and Credibility of Lessons Learned
- Usability of Lessons Learned

6. Criterion: Reporting
Viewpoints:
- Presentation/Legibility and Clarity
- Utilization of Tables and Figures
- Presentation of Primary Data

III. Project Evaluation Based on Written Report (DAC’s Five Evaluation Criteria)

1. Criterion: Relevance
Viewpoints:
- Validity
- Necessity
- Appropriate Approach

2. Criterion: Effectiveness
Viewpoints:
- Achievement Level of Project Purpose
- Causal Relationships between Outputs and Project Purpose

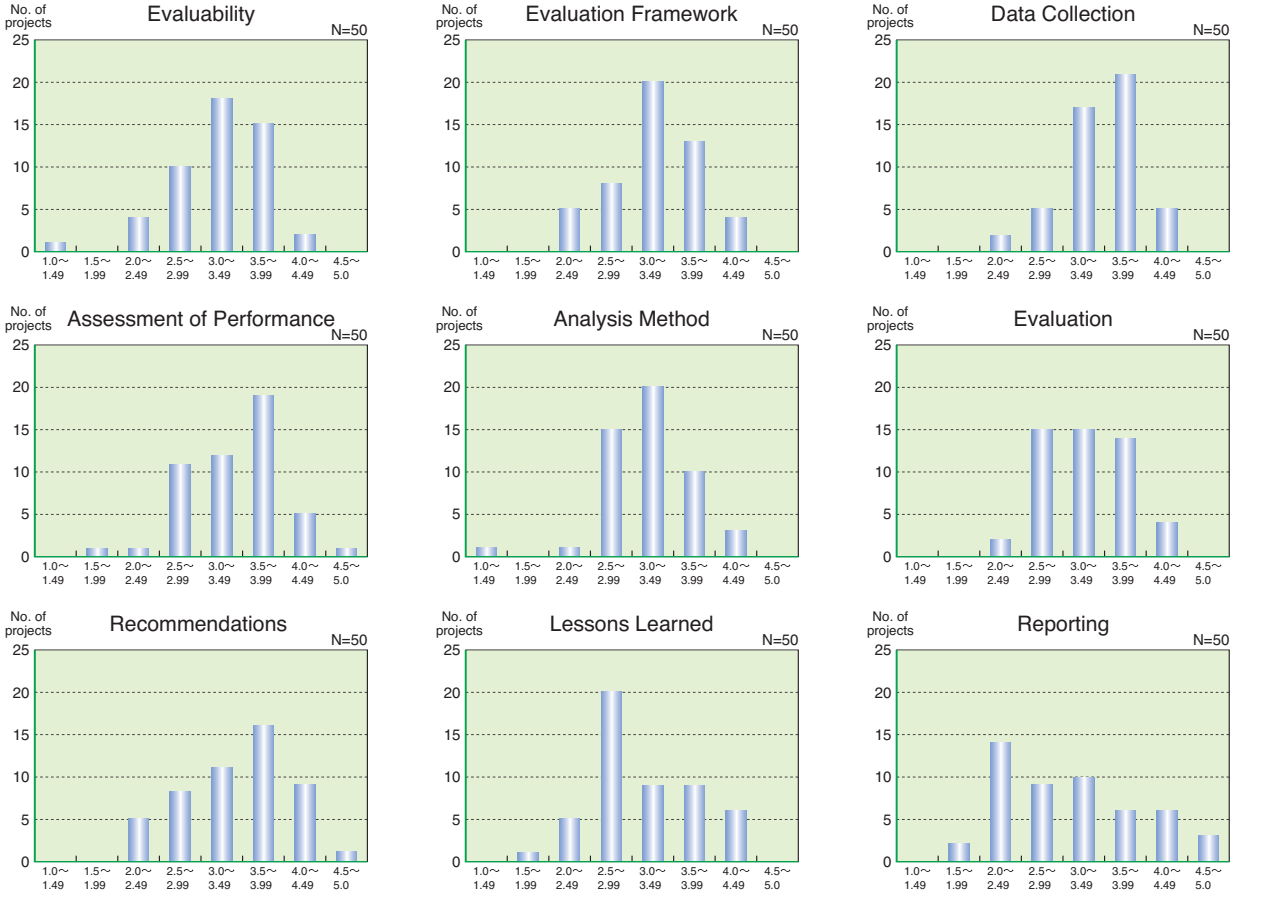
3. Criterion: Efficiency
Viewpoints:
- Clear Input Cost
- Cost-benefit Performance
- Appropriate Implementation Process

4. Criterion: Impact
Viewpoints:
- Achievement Level of Impact
- Logic on Causal Relationships of Impact
- Unanticipated Impact (Both Positive and Negative)

5. Criterion: Sustainability (Post-JICA’s Cooperation)
Viewpoints:
- Mechanism of Securing Sustainability
- Level of Sustainability
- Organizational Sustainability
- Technological Sustainability
- Financial Sustainability

6. Criterion: Overall Evaluation

Figure 4-3 Distribution of Scores by Evaluation Criterion



(2) Evaluation Results by Criterion

The viewpoints of each evaluation criterion were rated and qualitative data was collected in the form of evaluator comments written in additional boxes on the sheet. We will summarize the current conditions and issues of the quality of terminal evaluation by criterion based on the evaluation results of scores for the viewpoints of each evaluation criterion and comments from the evaluators. Figure 4-4 illustrates the average scores results for viewpoints under each evaluation criterion and those for evaluation criteria.

1) Evaluability

"Evaluability" is a criterion that questions the appropriateness of set conditions for an evaluation. This item is evaluated based on the following six viewpoints:

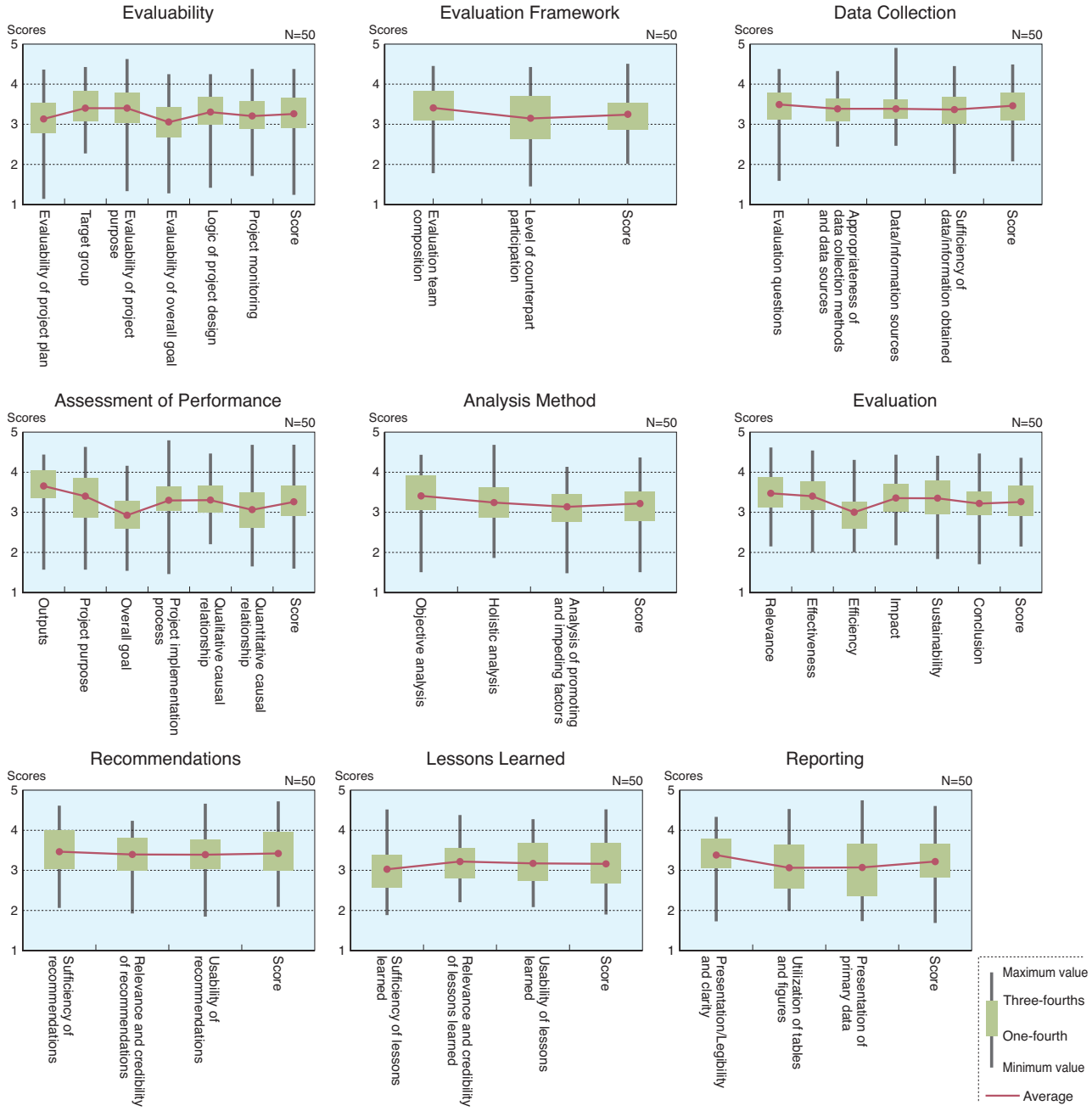
- Evaluability of project plan (preliminary study/PDM): whether the project plan (preliminary study/PDM) was appropriate for evaluating the project
- Target group: whether the target group or beneficiary of the project was set clearly and properly
- Evaluability of project purpose: whether the indicators and specific target values are clearly defined for each output and project purpose so that they can be used to measure the level of project achievement
- Evaluability of overall goal: whether the indicators and specific target values are clearly defined for overall goals so that they can be used to measure the level of project achievement
- Logic of project design: whether the PDM used for evaluation describes a clear and realistic logic flow from Overall goal - Project Purpose - Outputs - Inputs, in considering important external assumptions
- Project monitoring: whether outputs, activities, and inputs were regularly monitored, with information including statistical data accumulated during project implementation

The average scores are 3.0 or higher, securing the "medium" level or higher on the grading scale, but the quality of the evaluations varies in some viewpoints. For example, the scores for "target group," "evaluability of project purpose," and "logic of project design" are 3.3 or higher, securing the "medium" level or higher on the grading scale. Specifically, "target group" and "evaluability of project purpose" scored 3.4 on average, which is higher than that of other viewpoints. In contrast, "evaluability of overall goal" scored 3.0 on average, showing the lowest score of all viewpoints concerning "evaluability." The low score for "evaluability of overall goal" is due to the scarce causal relationships between overall

goal and project purpose, or because several few projects had vague overall goals and indicators.

High scores were given to (1) projects with specified indicators, clear criteria for project purpose, a well-established system for monitoring by the counterpart and beneficiaries, and thorough monitoring to accumulate data, (2) projects with target values for overall goal, project purpose, and outputs to easily evaluate achievement levels, and (3) projects with verifiable records obtained by a baseline survey. High scores were also given to projects that were highly verifiable, with a corrected PDM attached, together with each PDM clarifying the project purpose and outputs in a clear-cut time frame and target values, based on the explanation of how and why the initial PDM was corrected. Conversely, low scores were given to (1) projects for which no PDM was prepared at all, (2) projects for which a PDM was prepared only at the time of evaluation, (3) projects for which project purposes were vague or abstract, thereby making it hard to grasp achievement level, and (4) projects in which the target group was inappropriately set and quantitative indicators were available but no monitoring was conducted according to those indicators.

Figure 4-4 Score Results for Viewpoints under Each Evaluation Criterion (Average Scores and Distribution)



2) Evaluation Framework

"Evaluation framework" refers to the evaluators of terminal evaluation. This criterion consists of two viewpoints:

- Evaluation team composition: whether evaluation team members of the project are sufficiently qualified to conduct professional and impartial evaluations
- Level of counterpart participation: whether counterparts in the developing country participated sufficiently in the evaluation as evaluators

The average scores for these viewpoints all resulted in 3.3 points or higher, securing the "medium" level or higher. "Evaluation team composition" scored 3.5 on average, showing a higher score than those for viewpoints of other evaluation criteria. "Level of counterpart participation" scored 3.2 on average. The high score of "evaluation team composition" is considered due to the participation of counterparts in the evaluation team at the time of evaluation, and the evaluation team with expertise, impartiality, and fairness.

High scores were given to (1) projects in which the evaluation team members have high expertise, a certain number of members are involved, and the membership is well-balanced, and (2) projects in which the names and fields of evaluation team members from partner countries are specified, they accompany survey visits, and hold many joint evaluation meetings. In contrast, low scores were given to (1) projects for which reports contain no information about the evaluation team composition, (2) projects in which the partner countries don't take part in the evaluation, and (3) projects for which reports indicate that a joint evaluation has been conducted but fail to mention the names of evaluation members on the partner countries or how much the partner countries participated.

3) Data Collection

"Data collection" intends to check how data were collected. This criterion is assessed based on the following four viewpoints:

- Evaluation questions: whether specific and practical questions were properly set in line with the evaluation purposes, in order to contribute to realistic information collection plans
- Appropriateness of data collection methods and data sources: whether several different data collection methods were used to increase the objectivity and credibility of information obtained
- Data/information sources: whether the sources of data/information are explained adequately in the evaluation report
- Data collection: whether the data/information was sufficient to answer evaluation questions in terms of both quality and quantity

The viewpoints on "data collection" show little variance in the quality of evaluation, with all average scores higher than 3.4 and securing more than the "medium" level. In addition, "evaluation questions," which means that specific questions were properly set in line with evaluation purposes in the planning stage in order to contribute to realistic information collection plans, scored 3.5 on average. It shows a higher score than the viewpoints of other evaluation criteria.

High scores were given to (1) projects in which sufficient information is obtained from several obvious information sources by appropriate methods of collecting information, and (2) projects with sufficient information for evaluation collected by properly prepared evaluation questions and interviews conducted with other donors and final beneficiaries. Conversely, there were some projects in which information was improperly collected, such as (1) those with no evaluation grid, (2) those with biased sources of information based on the results of interviews with stakeholders within the project, (3) those with insufficiently grasping information concerning target values, and (4) those with unclear sources.

4) Assessment of Performance

"Assessment of performance" is evaluated with the following six viewpoints:

- Measurement of results (outputs): whether the achievement level of outputs was properly measured against the target values set by the indicators
- Measurement of results (project purpose): whether the achievement level of the project purpose was properly measured against the target values set by the indicators
- Measurement of results (overall goal): whether the achievement level of the overall goal was properly measured against the target values set by the indicators
- Examination of project implementation process: whether the project implementation process (monitoring, decision making, communication within the project) was thoroughly examined, through which impeding and/or promoting factors relative to achieving outputs, project purpose, and overall goal are identified
- Examination of qualitative causal relationships - logic of project design: whether the logic of project design was thoroughly verified, through which impeding and/or promoting factors relative to achieving outputs, project purpose, and overall goal are identified
- Examination of quantitative causal relationships - before and after: whether the causal relationships were thoroughly examined to verify that effects for beneficiaries have resulted from project interventions

The scores of each viewpoint for "assessment of performance" show great variance in evaluation quality. Of the viewpoints, "measurement of results (outputs)," "measurement of results (project purpose)," "examination of project implementation process," and "examination of qualitative causal relationships" all scored 3.3 or higher on average. "Measurement of results (outputs)," scoring 3.7 on average, has the highest score of all viewpoints. Conversely, "measurement of results (overall goal)" scored less than 3.0 on

average, falling short of the "medium" level and thus the lowest among viewpoints of all evaluation criteria. The assessment of performance of the overall goals also scored low in secondary evaluations in fiscal 2005 and 2006. And these low scores are considered due to some projects having vague overall goal and unclear indicators for those goals, resulting in failure or insufficiency in assessing performance.

High scores were given to (1) projects where data is collected through monitoring, the before/after approach is widely used, and overall goals, project purposes, and outputs are assessed quantitatively and qualitatively, and (2) projects where the achievement levels of outputs, project purposes, and overall goals are assessed based on the indicators and target values of the PDM. In contrast, low scores were given to (1) projects where achievement levels of their project purposes and overall goals are insufficiently verified, and (2) projects where verification is not conducted according to the indicators and target values.

5) Analysis Method

In “analysis method,” evaluation is judged based on the following three viewpoints to check how analysis is conducted:

- Objective analysis: whether the data was objectively analyzed based on a series of scientific discussions, with an effort made to quantify the data where feasible
- Holistic analysis: whether data interpretation was based on an examination and analysis of different methods, and from various aspects
- Analysis of promoting and impeding factors: whether factors that promote and impede effects were adequately analyzed in light of the project logic (cause-effect) and project implementation process (such as project management)

The average score results for all viewpoints achieved 3.1 or higher (average) and attained the “medium” level. "Analysis of promoting and impeding factors" scored relatively lower on average than the scores for other viewpoints.

High scores are given to (1) projects that collected qualitative and quantitative data from various information sources and conducted holistic and detailed analyses objectively, and (2) projects that used qualitative and quantitative data to analyze the promoting and impeding factors regarding the outcome of the project. Conversely, low scores were given to (1) projects that conducted interviews, but failed to clarify whether their judgments were made based on those findings and did not conduct analyses based on data, and (2) projects that did not collect sufficient and objective information, thus failing to give clear grounds for activity results.

6) Evaluation

“Evaluation” means to evaluate based on six evaluation viewpoints: DAC's five evaluation criteria plus “conclusion” used to check whether the conclusion was drawn based on holistic viewpoints relative to the five evaluation criteria.

- Relevance: whether perspectives for evaluating "Relevance” (e.g., validity and necessity of a project in view of beneficiaries' needs, project implementation as an appropriate approach to problem solving, consistency of policies) were sufficiently covered
- Effectiveness: whether perspectives for evaluating "Effectiveness” (e.g., achievement level of project purpose, causal relationships between outputs and project purpose) were sufficiently covered
- Efficiency: whether perspectives for evaluating "Efficiency” (e.g., comparison with other similar projects through cost analysis, cost-effectiveness analysis) were sufficiently covered
- Impact: whether perspectives for evaluating “Impact” (e.g., achievement level of overall goal, causal relationships between project purpose and overall goal) were sufficiently covered
- Sustainability: whether the perspectives for evaluating “Sustainability” (e.g., mechanism of securing sustainability and outcomes to be produced in terms of policies and systems, organizational and financial aspects, technical aspects, socio-culture, environment) were sufficiently covered
- Conclusion: whether the conclusion was drawn based on holistic viewpoints relative to the five evaluation criteria

The average score results for all viewpoints were 3.0 or higher, exceeding the “medium” level, yet showing a variance in evaluation of the viewpoints. Among the viewpoints, the average score for “relevance” is the highest at 3.5 and relatively higher than viewpoints in other evaluation criteria. As for “efficiency,” the average score was the lowest of the six viewpoints at 3.0 and relatively lower than the viewpoints in other evaluation criteria. “Efficiency” was scored the lowest in the secondary evaluations in fiscal 2005 and 2006 as well. This criterion questions whether perspectives (e.g., comparison with other similar projects through cost analysis, cost effectiveness) are sufficiently covered. Many projects did not mention a comparison with similar projects and cost-effectiveness, and failed to provide adequate information on actual expenses, thus leading to the low score.

High scores were given to (1) projects that fully explained the DAC's five evaluation criteria, gave an appropriate conclusion, and elaborated on the analysis of promoting and impeding factors, and (2) projects that properly analyzed each item in line with a particular viewpoint and described an abstract of each evaluation together with the grounds thereof to be understandable. Conversely, low scores were given to (1) projects where an evaluation with clear evidence according to viewpoints was not conducted and its conclusion was missing, and (2) projects that collected insufficient data required for evaluation and had no fixed viewpoint of evaluation, thereby causing an inadequate evaluation.

7) Recommendations

The criterion of “recommendations” concerns the following three viewpoints:

- Sufficiency of recommendations: whether the recommendations fully consider all promoting and impeding factors identified during the evaluation process
- Relevance and credibility of recommendations: whether the recommendations are based on information obtained through the process of data analysis and interpretation, thereby resulting in objective and convincing recommendations
- Usability of recommendations: whether recommendations are practical and useful for feedback and follow-ups, within a specific time frame

The average scores for these viewpoints are 3.4 or higher, securing a relatively high evaluation considering the small variance in quality in the evaluations. Among the viewpoints, the average score for “sufficiency of recommendations” is 3.5 and relatively higher than scores for viewpoints in the other evaluation criteria.

High scores were given to (1) projects where concrete and highly practical recommendations were stated based on the current situation, and (2) projects where recommendations were distinguished between before and after project termination based on evaluation analysis so that they would be useful. Conversely, low scores were given to (1) projects with insufficient recommendations in which analysis results were not reflected, and (2) projects with unclear and non-specific recommendations and thereby resulting in low usability.

8) Lessons Learned

The criterion of “lessons learned” includes the following three viewpoints:

- Sufficiency of lessons learned: whether the lessons learned fully consider the impeding/promoting factors identified during the evaluation process
- Relevance and credibility of lessons learned: whether the lessons learned are based on information obtained through the process of data analysis and interpretation, thereby resulting in objective and convincing lessons learned
- Usability of lessons learned: whether the lessons are generalized and conceptualized for wide application

The average scores are 3.1 or higher, securing the “medium” level or higher. The average score for “sufficiency of lessons learned” is 3.1 and relatively lower than the other two viewpoints.

High scores were given to projects that gave specific and useful lessons learned based on evaluation analysis, such as the effects of the project approach and collaboration with other donors, and to projects that gave highly usable lessons learned by reviewing the problems drawn from the analysis of impeding factors, on the premise of application to similar projects. In contrast, low scores were given to projects where the specific lessons learned were inadequately considered so that they were poorly applicable to similar projects, or projects where the issues to be solved were described as lessons learned.

9) Reporting

“Reporting” covers the following three viewpoints:

- Presentation/legibility and clarity: whether the evaluation report is clear and simple, and thus understandable to readers in terms of structure, font, terminology, and data presentation
- Utilization of tables and figures: whether tables and figures are effectively utilized to visually present statistics and analysis results
- Presentation of primary data: how sufficient primary data such as on targets, interview and questionnaire results, or sources are presented properly in the report

Every viewpoint in "Reporting" scored 3.1 or higher on average and thus ranks higher than the "medium" level. "Presentation/legibility and clarity" scored 3.4 on average and is thus higher in quality of evaluation than the other viewpoints.

High scores were given to (1) projects that gave evaluations according to the PDM, effectively used many flowcharts and other visuals, were rich in primary data, and written in a brief and logical manner, thereby easy to understand, and (2) projects that presented questionnaire questions and qualitative responses, and gave persuasive descriptions, making good use of tables and figures. Conversely, low scores were given to (1) projects that failed to specify actual activities or results to cover the items necessary for the report, and (2) projects that lacked primary data, tables and figures, and are written in a redundant manner.

(3) Examples of Good Quality Evaluation Reports and Poor Quality Evaluation Reports

The revised JICA Guidelines for Project Evaluation (March 2004) explains in detail important points to be considered for appropriate evaluation. However, it is not easy to write a report that is easy to understand and highly qualified. If some reports of terminal evaluations that are highly qualified are presented using the results of secondary evaluation, the reports of these evaluations can serve as role models. And if evaluation studies and reporting are conducted with reference to the methods and contents in these models, the quality of reports will be secured more easily.

The quality of terminal evaluations was evaluated from nine evaluation criteria: “evaluability,” “evaluation framework,” “data collection,” “assessment of performance,” “analysis method,” “evaluation,” “recommendations,” “lessons learned,” and “reporting.” The overall quality of terminal evaluations was judged based on the value obtained by dividing the total scores for the nine evaluation criteria by the number of criteria. The highest attainable score is five; the lowest is one, and “medium” level is three. We selected the

top five cases of evaluations (including two fourth-ranked cases) and the worst four cases, while giving consideration to the distribution of overall scores. The web graphs in Figures 4-5 and 4-6 respectively show the scores for the nine criteria of these eight evaluations. Table 4-3 lists the average scores and differences in average scores for evaluation criteria of the top five and worst four evaluations.

As clearly seen in Figure 4-5 and Table 4-3, the average scores of the top five evaluations are quite high at 4.0 or higher for “data collection,” “assessment of performance,” “analysis method,” “evaluation,” “lessons learned,” and “reporting.” In particular, “reporting” scored very high at 4.4. In specific terms, the factors contributing to such high scores are (1) the sufficient collection of appropriate data from clear data sources, (2) the full assessment and examination of the implementation process, performance and effects of the projects from qualitative and quantitative aspects, (3) the objective analysis of collected data from various aspects, (4) analysis of promoting and impeding factors for producing outcomes, (5) evaluations covering the necessary prospective, (6) the drawing of objective and convincing lessons useful for similar projects from information on the promoting and impeding factors obtained during the evaluation processes, and (7) well-prepared reports which are easy to understand.

As evidenced in Figure 4-6 and Table 4-3, there is conversely a tendency among the worst four projects where the scores for “evaluability,” “analysis method,” “lessons learned,” and “reporting” are relatively low. Vague and illogical indicators for project purposes and overall goals result in insufficient quantitative analysis, thereby failing to produce persuasive analysis results. Moreover, the lessons learned are insufficiently identified, and that causes a presentation of quite usual lessons. The report doesn't indicate primary data or what is necessary. These factors lead to low scores regarding the quality of terminal evaluation. Moreover, projects with low overall scores clearly show low scores for all criteria.

In every evaluation criterion, the difference in average scores between the top five and worst four projects is statistically significant, with the top five projects rated higher than the worst four. The difference in average scores for both groups is smallest in "evaluation framework" and largest in "reporting." A large difference was found between the two groups in the average scores for “evaluability,” “data collection,” “assessment of performance,” “analysis method,” and “lessons learned,” indicating that these criteria are important factors for evaluating the quality of reports.

Table 4-4 lists the five evaluation reports rated as being high in quality; Table 4-5 lists the four evaluation reports rated as being poor in quality.

The “Project for the Capacity Building of the National Institute of Occupational Safety and Health” in Malaysia obtained the highest score for its terminal evaluation. Malaysia is achieving high growth but undergoing an upward trend toward heavy occupational hazards and work-related

Figure 4-5 Top 5 Projects for Quality of Terminal Evaluation

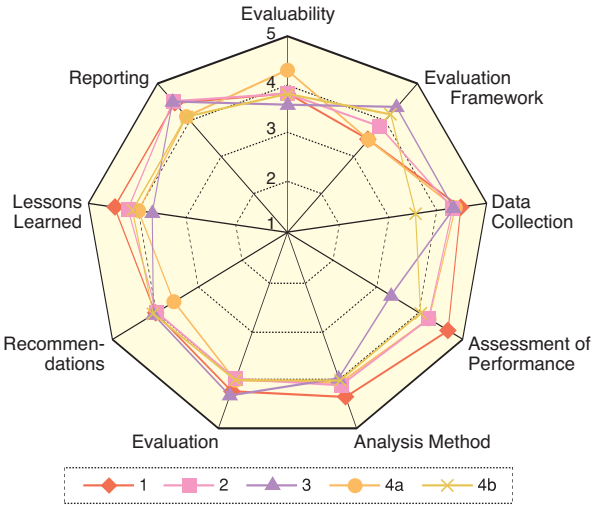


Figure 4-6 Worst 4 Projects for Quality of Terminal Evaluation

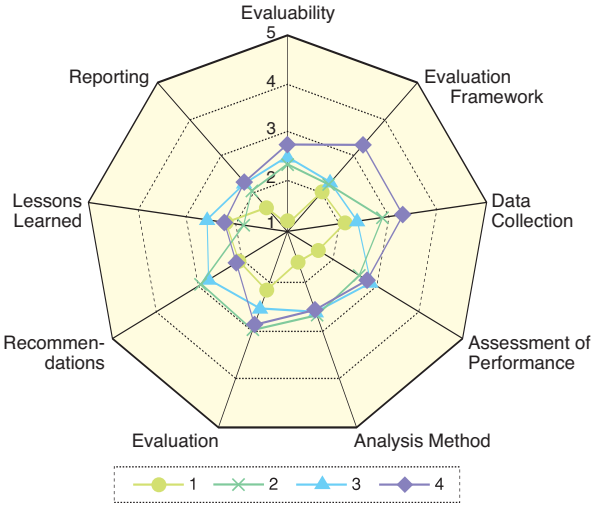


Table 4-3 Score Results of Top 5 and Worst 4 Projects (Average)

Evaluation Criteria	Average Scores		Difference in Average Scores
	Top 5 projects	Worst 4 projects	
Evaluability	3.92	2.29	1.63 **
Evaluation Framework	3.90	2.52	1.38 **
Data Collection	4.19	2.67	1.52 **
Assessment of Performance	4.11	2.47	1.64 **
Analysis Method	4.08	2.33	1.75 **
Evaluation	4.08	2.63	1.45 **
Recommendations	3.92	2.50	1.42 **
Lessons Learned	4.10	2.29	1.80 **
Reporting	4.36	2.15	2.21 **

** The difference in significance level between the top five and worst four projects is 1% on average.

diseases. To improve occupational safety and health in governmental organs, private companies, factories, and other entities, a technical cooperation project was implemented for the capacity building of the National Institute of Occupational Safety and Health. The transfer of technology progressed successfully, with mobile physical examination service being provided for the first time in Malaysia.

With regard to the quality of terminal evaluation, the evaluation criteria of “data collection,” “assessment of performance,” “analysis method,” “evaluation,” “lessons learned,” and “reporting” all scored 4.2 or more. The terminal evaluation of this project was deemed very qualified for the following reasons: Questions in the evaluation grid are appropriate. Cross-checking is made for each question with information obtained from several sources. Sufficient information necessary for evaluation has been collected, showing the sources clearly. The before/after approach was employed to quantitatively and qualitatively assess the project. Moreover, objective analysis from various aspects has been conducted including an analysis of promoting and impeding factors. The evaluation is conducted fairly based on the analysis findings, with logical conclusions made according to the process. The recommendations cover the direction that the National Institute of Occupational Safety and Health should take after project completion. The contents are appropriate, with the lessons learned both concrete and useful. The report is well organized, written in clear language, and uses many organized and legible tables.

The “Cross Border Initiative Project” in Zambia is the worst project regarding the quality of its terminal evaluation. This project was launched to succeed the community empowerment program jointly implemented by Japan and the USA. Continuing the framework of Japan-U.S. collaboration, USAID contributed funding to an NGO in the USA, while JICA implemented technical cooperation to an NGO in Zambia. The project aims to prevent HIV transmission among high risk groups in the border region. The project activities enabled high-risk groups to receive continuous treatment without discrimination. From 2000 to 2006, the sexually transmitted infection rate of sex workers' partners has been slightly decreased. Thus, it is assumed that knowledge of sexually transmitted infection and HIV among sex workers has been upgraded. It remains unclear, however, whether this result is due to the project.

The quality of terminal evaluation indicates that scores of "evaluability," "assessment of performance," "analysis method," and "reporting" are in the 1.0 range, while other evaluation criteria score less than 2.3. The quality of terminal evaluation is considered low for the following reasons: “Assessment of performance” is evaluated based on data collected through interviews with NGO insiders. Few data that indicate the performance objectively are collected. Evaluations are not carried out in line with the evaluation questions. The evaluations of "effectiveness" and "impact" are based on unpersuasive grounds, and a concrete description of the output of the activities is not presented, thus failing to be clear. As for "recommendations" and "lessons learned," only the issues to be solved are described without showing anything concrete. The actual activities or performance are not described concretely in the report. Therefore, it is impossible to grasp the activities of the project.

(4) Year-to-Year Changes in the Quality of Evaluation

The secondary evaluation of terminal evaluations has been conducted since fiscal 2003, targeting 38 terminal evaluations conducted in fiscal 2002, 38 in fiscal 2003, 45 in fiscal 2004, 42 in fiscal 2005, and 25 in fiscal 2006, besides those conducted this fiscal year. The project evaluations should guarantee a high quality in order to accurately convey the outcomes of implemented projects to readers. We will examine how the quality of terminal evaluation has changed over the years with focus on the projects subject to secondary evaluation in fiscal 2003 and afterwards, of which evaluation criteria and viewpoints were similar.

The evaluators of secondary evaluation are different every year. Even though the evaluation criteria remain the same, evaluation viewpoints have differed slightly over the years. Rating scales have also changed. A ten-level rating scale was adopted in fiscal 2004, whereas a five-level rating scale was used in fiscal 2005, 2006, and 2007. Thus, considering the differences in evaluation standards, the evaluation results of projects evaluated twice were used to convert the evaluation scores to unify the evaluation scale. In other words, since the

Table 4-4 Terminal Evaluations of Good Quality

	Country	Project Name	Total Score	Fiscal Year of Evaluation
1	Malaysia	Project for the Capacity Building of the National Institute of Occupational Safety and Health	4.22	2005
2	Viet Nam	Forest Fire Rehabilitation Project	4.11	2006
3	Niger	Project on Support to the Improvement of School Management through Community Participation ("School For All")	4.02	2006
4a	Senegal	Project on the Safe Water and the Support on Community Activities	4.01	2005
4b	China	Model Planning Project for Water-saving Measures on Large-scale Irrigation Scheme	4.01	2005

Table 4-5 Terminal Evaluations of Poor Quality

	Country	Project Name	Total Score	Fiscal Year of Evaluation
1	Zambia	Cross Border Initiative Project	1.88	2005
2	Indonesia	Technical Cooperation for Community Empowerment Program with Civil Society	2.54	2006
3	Micronesia	Fisheries Training Project (Extended)	2.58	2005
4	Costa Rica	Project on Productivity Improvement for Enterprises	2.72	2005

evaluation scores from fiscal 2003 to 2005 have already been converted to the scale used in fiscal 2006 in the previous report, these scores were converted to the scale in fiscal 2007. It must be noted that the terminal evaluations subject to secondary evaluation twice adopted the original scores, not the converted ones.

In fiscal 2004 and 2005, the recommendations and lessons learned were also lumped together as a single criterion, while in fiscal 2006 and 2007, they were separated into respective criteria ("recommendations" and "lessons learned"). Thus, for the results of secondary evaluations conducted in fiscal 2004 and 2005, both the total scores of three viewpoints on "recommendations" and the total scores of three viewpoints on "lessons learned" were divided proportionally according to respective percentages of the total score in an attempt to evaluate "recommendations" and "lessons learned" separately.

The average scores of 38 projects in fiscal 2003, 45 projects in fiscal 2004, 42 projects in fiscal 2005, and 25 projects in fiscal 2006 were obtained by evaluation criterion, as shown in Figure 4-7 and listed in Table 4-6. As for changes in average scores between 2003 and 2004, the scores in 2004 are significantly higher statistically for "evaluation," "recommendations," and "reporting." Regarding the differences between fiscal 2003 and 2005, the average scores for "evaluation," "recommendations," "lessons learned," and "reporting" in fiscal 2005 are significantly higher statistically. Comparing fiscal 2003 and 2006, the average scores for "evaluation" and "reporting" in fiscal 2006 are statistically higher. The average scores in fiscal 2004 and afterwards tend to be high in general compared to the terminal evaluation in fiscal 2003,.

Secondary evaluation in fiscal 2006 concluded that the quality of terminal evaluations in fiscal 2004 and 2005 was higher than that in fiscal 2003. Seventeen projects in fiscal 2005 were subject to secondary evaluation in fiscal 2006, while projects targeted for secondary evaluation in fiscal 2007 were increased to 42. Thus, the results of secondary evaluation in 2007 are more stable. Though only 25 projects in fiscal 2006 were targeted for secondary evaluation in fiscal 2007, the quality of terminal evaluation in fiscal 2006 was also upgraded compared to that in fiscal 2003. Based on these results, it can be concluded that the quality of terminal evaluation has been improved and maintained since fiscal 2004. The difference of average scores between fiscal 2004, 2005 and 2006 is not statistically significant.

(5) Evaluation by JICA Headquarters and Overseas Offices

In line with the decentralization of operations, JICA started "projects in overseas offices in charge" at eight overseas offices on a trial basis since October 2004, and has since been expanded to 30 offices in 2005. Under this system, the overseas offices are entrusted with the authority to carry out a series of operations from project formulation to implementation and project evaluation. Likewise, terminal evaluations that had been conventionally conducted by the headquarters were gradually taken over by overseas offices for these projects under direct management. Among the target projects for secondary evaluation, two in fiscal 2005 and six in fiscal 2006 were conducted by overseas offices.

Figure 4-7 Year-to-Year Changes in the Quality of Evaluations (Average Scores)

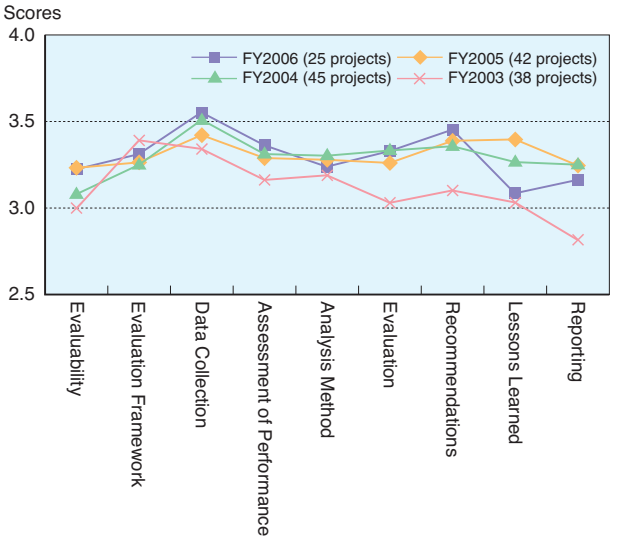
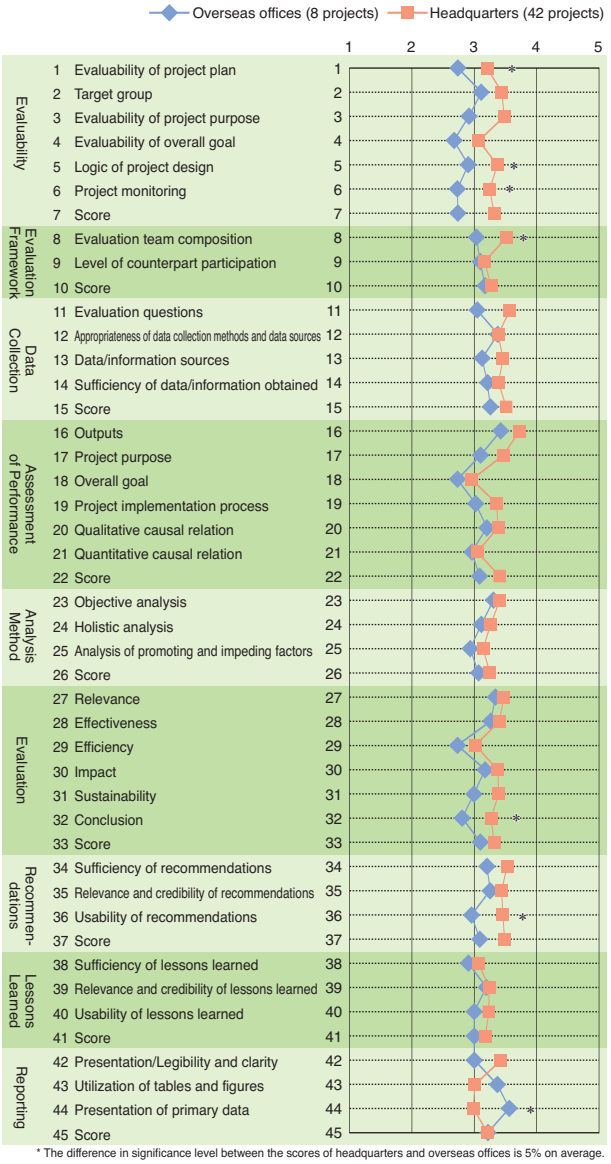


Table 4-6 Year-to-Year Changes in the Quality of Evaluations (Average Scores)

	Average Scores				Difference in Average between Years					
Evaluation Criteria	(1) 2003	(2) 2004	(3) 2005	(4) 2006	(2)–(1)	(3)–(1)	(4)–(1)	(3)–(2)	(4)–(2)	(4)–(3)
I Preconditions for Conducting Appropriate Evaluation										
Evaluability	3.01	3.08	3.23	3.22	0.08	0.23	0.22	0.15	0.14	– 0.01
II Key Evaluation Criteria										
Evaluation Framework	3.40	3.26	3.27	3.31	– 0.14	– 0.13	– 0.09	0.01	0.05	0.04
Data Collection	3.35	3.51	3.43	3.55	0.16	0.08	0.20	– 0.09	0.04	0.13
Assessment of Performance	3.17	3.32	3.29	3.37	0.15	0.13	0.20	– 0.02	0.05	0.07
Analysis Method	3.20	3.31	3.28	3.24	0.11	0.08	0.04	– 0.03	– 0.07	– 0.04
Evaluation	3.03	3.33	3.26	3.34	0.30 **	0.23 *	0.30 **	– 0.07	0.00	0.07
Recommendations	3.11	3.36	3.39	3.46	0.25 *	0.28 *	0.35	0.03	0.10	0.06
Lessons Learned	3.04	3.27	3.40	3.09	0.23	0.36 *	0.05	0.13	– 0.18	– 0.31
Reporting	2.82	3.25	3.26	3.17	0.43 **	0.44 **	0.35 *	0.00	– 0.09	– 0.09

* The difference in significance level between the scores in fiscal years is 5% on average.
** The difference in significance level between the scores in fiscal years is 1% on average.

Figure 4-8 Quality of Evaluations Conducted by Headquarters and Overseas Offices



* The difference in significance level between the scores of headquarters and overseas offices is 5% on average.

Figure 4-8 shows the differences in scores between 42 projects subjected to terminal evaluation by headquarters and eight projects subjected to terminal evaluation by overseas offices. Though there are constraints as overseas offices conducted terminal evaluation only for eight projects, terminal evaluation reports by overseas offices reveal scores of 3.0 or higher in many evaluation criteria which means "medium" level or higher according to this figure. When comparing the terminal evaluation by headquarters with that by overseas offices, however, the average scores in many evaluation criteria of terminal evaluation by overseas offices tend to be lower than those by headquarters. Statistically significant differences are revealed in the viewpoints of "evaluability of project plan," "logic of project design," and "project monitoring" in "evaluability." Therefore, the scores of overseas offices are lower than those of headquarters, and the average scores of overseas offices are mostly in the 2.0 range. "Evaluation team composition" in "evaluation framework" shows a statistically significant difference, and score of overseas offices is lower than that of headquarters. As for "evaluation," "conclusion" shows a statistically significant difference, and the score of overseas offices is lower than that of headquarters. However, with respect to two viewpoints of "utilizations of tables and figures" and "presentation of primary data" in "reporting," overseas offices achieved higher scores than headquarters, with a statistically significant difference in "presentation of primary data."

Projects under overseas offices in charge comprise projects formulated by overseas offices and those by headquarters. As for the quality of terminal evaluation reports conducted by overseas offices, no major difference was found between projects formulated by headquarters and those by overseas offices. However, although there are constraints as only two projects were formulated and evaluated by overseas offices, the quality of these reports was generally low. This is presumably one of the reasons why the reports by overseas offices were lower in quality than those by headquarters.

(6) Implementation of Ex-ante Evaluation

JICA established a consistent evaluation system from ex-ante to ex-post throughout the implementation cycle of a project. As part of such efforts, ex-ante evaluation was introduced in fiscal 2001 to examine the necessity and priority of a project as well as the appropriateness of a project plan based on the expected effects prior to launching the project.

Among the projects targeted for secondary evaluation in fiscal 2007, nine were subjected to ex-ante evaluation in fiscal 2005, followed by 15 in fiscal 2006.

Whether an ex-ante evaluation was conducted is assumed to significantly affect the quality of the project and evaluation report. From the viewpoint of conducting a consistent evaluation from ex-ante to terminal, the project should set its measurable targets for each stage beforehand, from the same viewpoint to be measured for the terminal stage. Also, the project should sufficiently examine the causal relationship between the project purpose and overall goal, collect information on the initial state of indicators as well, and indicate well-grounded estimates of indicators for changes after the project ends. As such, if the project excels in the logic of the causal relationships of various results and sets the necessary indicators as well, the project itself will be well implemented. The evaluation report is also expected to be a convincing and high-quality one if written according to the changes in various indicators.

To analyse the relation between ex-ante and terminal evaluations, we compared the average scores between projects with and without ex-ante evaluation. The results exhibited no significant diference statistically in the two groups for all the evaluation criteria. However, since the average scores of projects with ex-ante evaluation were somewhat higher than those without it in many evaluation criteria, it is expected that ex-ante evaluation will improve the quality of terminal evaluation reports.

(7) Summary of the Quality of Primary Evaluation

It can be concluded that terminal evaluation meets a certain level of quality. The secondary evaluation in fiscal 2006 showed a relatively low score for quality regarding "evaluation framework." But the secondary evaluation in fiscal 2007 obtained the "medium" level or higher, and quality of evaluation is similar to that in other evaluation criteria. "Evaluation framework" is evaluated from the viewpoints of "evaluation team composition" and "level of counterpart participation." Both viewpoints show evaluation quality of the "medium" level or higher. Of the reports targeted this time, more reports now indicate the names and expertise of counterparts than in previous projects, thereby making it easier for secondary evaluators to obtain information about each viewpoint. This is presumably one of the reasons for the improved quality of evaluation. However, there still exist some projects where terminal evaluation was conducted only by the Japanese evaluation team, with the results finalized by the partner country without sufficient time. To increase participation of the partner countries in evaluation and ensure the expertise and impartiality of evaluators, the report must include involvement of the partner counry in the evaluation process in order to improve the quality of evaluation.

The quality of terminal evaluation is generally of the "medium" level or higher, but evaluation from the viewpoints in each evaluation criterion reveals some differences in quality. Specifically, the viewpoints of "evaluability of overall goal" in "evaluability," "overall goals" in "assessment of performance," and "efficiency" in "evaluation" show a lower quality of evaluation than other viewpoints. The evaluation check sheet is designed to have a relation between the viewpoints of "evaluability" and "assessment of performance." As for "evaluability of overall goal" in "evaluability," unclear indicators for overall goal and a vague causal relationship between the project purpose and overall goal when designing PDM presumably affects the primary evaluation of overall goals in assessment of performance. Therefore, clarifying the indicators for overall goals and the causal relationship between project purpose and overall goal will presumably help assess "evaluability" and "assessment of performance" of overall goals.

In terms of "efficiency" in evaluation, many projects are evaluated for their efficiency from the viewpoint of the implementation process, such as dispatch period of experts and the utilization of equipment provided. Few evaluations were conducted from the viewpoint of cost-effectiveness, such as cost comparison with similar projects. Therefore, evaluating efficiency from the viewpoint of whether the project has produced outcomes that match the input cost would be necessary to ensure accountability to the public.

The secondary evaluation in fiscal 2006 also revealed a relatively low trend in evaluation quality for "evaluability of overall goal" in "evaluability," and "overall goals" in "assessment of performance," and "efficiency" in "evaluation." For that reason, preparing an appropriate PDM when implementing a project and conducting primary evaluation based on cost-effectiveness are necessary to improve the quality of evaluation.

In "reporting," the evaluation on the utilization of tables and figures, and presentation of primary data such as survey results reaches the "medium" level, but is not so high when compared to the viewpoints of other evaluation criteria. Terminal evaluation reports should be written in an objective manner by clearly indicating survey results and in a legible manner by utilizing tables and figures, since a terminal evaluation report is open not only to people related to the project but also to the public.

Regarding the chronological changes in the quality of terminal evaluation, the quality of terminal evaluation in and after fiscal 2004 is higher than in 2003 and maintained. However, no clear changes are seen over time between fiscal 2004, 2005, and 2006.

JICA adopted “overseas office's direct project management” in 2004 by granting authority to conduct a series of operations for projects. Overseas offices have conducted terminal evaluations of these projects. Comparing the terminal evaluation by overseas offices with that by JICA headquarters, "evaluability of project plan," "logic of project design," and "project monitoring" in "evaluability," and "evaluation team composition" in "evaluation framework" show lower quality in primary evaluation by overseas offices than that by headquarters, thus indicating a statistically significant difference, although some criteria in primary evaluation by overseas offices is higher in quality. Logical monitoring in project evaluation is weaker, and assessment and verification of performance are insufficient. Moreover, evaluation was not conducted in line with JICA's evaluation guidelines in some projects. These are presumably the reasons for the lower quality of terminal evaluation reports by overseas offices. It is presumed that an increasing number of evaluations are conducted under the leadership of overseas offices in the future. Evaluations of high quality are made possible when the headquarters takes measures for qualitative improvement in evaluations implemented by overseas offices, through such means as guidance regarding evaluation methods and a thorough implementation of evaluation standards, and when the overseas offices conduct evaluations in line with JICA's guidelines such as preparation of an appropriate PDM.

(8) Improving Primary Evaluation

As described above, secondary evaluation was conducted based on the terminal evaluation results. The quality of primary evaluation is gradually improving as more reports refer to participation of the partner country in "evaluation framework" when reading the terminal evaluation report on secondary evaluation in fiscal 2007. However, similar factors seen in fiscal 2005 and 2006 to improve the quality of evaluation are also identified in the analysis results and comments by secondary evaluators. The following describes the key factors to improve the quality of primary evaluation.

1) Evaluation Team Composition

To improve the quality of evaluation, the partner countries must participate more in the evaluation. It is also necessary to indicate in the report how the counterparts were involved in the project and the extent of counterpart participation in joint evaluation (e.g. the Japanese side conducted the survey and only discussion was held jointly, or the counterpart participated in the entire evaluation process as a team member) so that the expertise and impartiality of evaluators are ensured.

2) Designing an Appropriate PDM

When a project is implemented, a PDM should have been designed. The evaluation is therefore conducted according to the PDM. It is doubtful in some projects whether its PDM was designed at all as some reports had no PDM attached and presented no evaluation grid. Some reports also used unclear indicators as targets or had a vague causal relationship between project purposes and overall goals, thereby failing to assess performance sufficiently. Since the setting of project purposes and overall goals is considered to affect primary evaluation regarding the overall goals in assessment of performance, it is necessary to clearly determine the causal relationship among outputs, project purposes and overall goals, along with the various indicators.

3) Data Collection

Some reports were evaluated as being low in quality due to an insufficient assessment of performance resulting from the insufficient collection of data. When the indicators used to measure the achievement level of project purpose cannot be obtained by the means specified in the PDM, it becomes necessary to employ an alternative means of obtaining information.

In some cases of data collection, the sources were restricted to a limited range of people. Even for short-term surveys, it is necessary to widen the sources of data collection so that the objectivity of evaluation is secured.

Few reports analyzed efficiency from the viewpoint of cost-effectiveness. Evaluation from the viewpoint of whether a project has achieved the outcomes that match the input cost is necessary to fulfill its accountability to the public. It is therefore necessary to collect data that enables a cost comparison with similar projects.

4) Objective Analysis

Some reports made judgments not based on their own survey results, but based on conjecture from the survey results of another establishment. There's also a gap between the conclusion and survey results in some reports. To ensure objectivity, it is necessary to clarify the progress and grounds for conclusions.

5) Recommendations and Lessons Learned

If the recommendations are general and abstract, what should be done to achieve the project purposes is unclear. It is necessary to fully identify the impeding factors and other elements in the implementation process, and describe necessary measures more concrete so that they are easily utilized. It is recommended that the lessons learned specified from the promoting and impeding factors are described concretely, by considering whether the lessons learned are applicable to similar projects.

6) Reporting

Terminal evaluation reports should be written clearly by indicating the survey results to secure the objectivity and utilizing the tables and figures for easy understanding by the readers, considering that the reports are open to not only project stakeholders but also the public.

7) Guidance on the Evaluation Methods for Overseas Offices

Secondary evaluators evaluated the terminal evaluation reports made by overseas offices with lower scores than those made by headquarters. It is presumed that an increasing number of evaluations will be conducted under the leadership of overseas offices in the future. Headquarters will set up measures intending qualitative improvement in evaluations implemented by overseas offices, through guidance regarding evaluation methods and a thorough implementation of evaluation standards. Moreover, overseas offices should always conduct evaluation according to JICA's guidelines such as preparing an appropriate PDM.

3. Project Evaluation by Secondary Evaluators Based on Terminal Evaluation Reports

(1) Summary of the Secondary Evaluation of Projects

We conducted a secondary evaluation of 25 projects evaluated in fiscal 2005 and 25 projects in fiscal 2006 by using terminal evaluation reports from the perspective of the DAC's five evaluation criteria. Figure 4-9 shows the results of project evaluation gleaned from the reports made by secondary evaluators.

All average scores for the projects are in the 3-point range, and thus above the "medium" level. Among the five evaluation criteria, the average score for "relevance" was the highest at 3.8 points, while the average score for "efficiency" was relatively low at less than 3.1 points.

Figure 4-10 shows the distribution of scores for project evaluation. All scores for "relevance" are higher than 2.5 points, including some at 4.5 points or higher. Most of the scores are generally clustered between 3.5 and 3.99 points in the higher range. The scores for "effectiveness" are mostly 3.0 points or higher, with some being in the 2-point range and some at 4.5 points or higher. No more than 4.5 points are given to "efficiency," and most scores fall within the range of 2.5 and 3.49, showing relatively low ratings. The scores regarding "impact" scatter not in the 1-point range or 3.5 points or above, with most falling between 3.0 and 3.99. Most scores for "sustainability" are between 2.0 and 3.99, with a large variance among the projects; some are in the 1-point range but none reached 4.5 points or higher, thus reflecting relatively low ratings.

(2) Project Evaluation from Viewpoints for Each Criterion

We conducted secondary evaluation from various viewpoints for the five evaluation criteria based on information obtained from the reports. Figure 4-11 shows the average scores and distributions for the viewpoints.

1) Relevance

The average scores for all viewpoints of "relevance" are high. Among the three viewpoints, the average scores for "validity" and "necessity" of project implementation are relatively high at more than 3.8 points. The average score for "appropriate approach" regarding whether the approach was appropriate and effective for achieving project purposes is 3.2 points, which is relatively lower than those for other viewpoints.

Figure 4-9 Project Evaluation by Secondary Evaluators

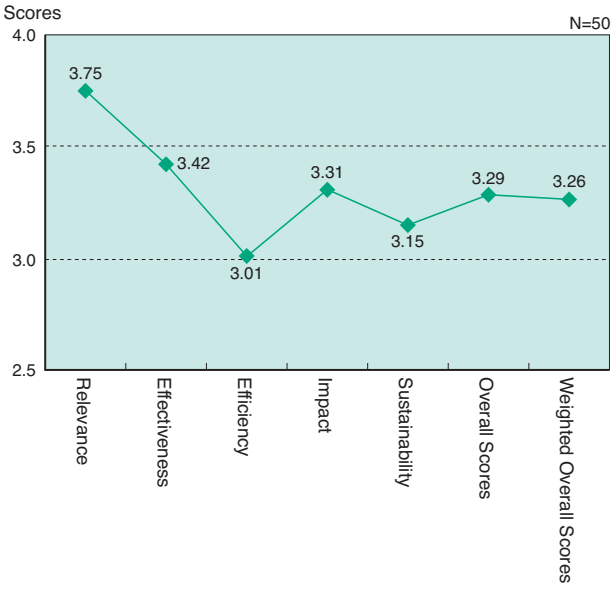


Figure 4-10 Distribution of Evaluation Scores for Projects by Secondary Evaluators

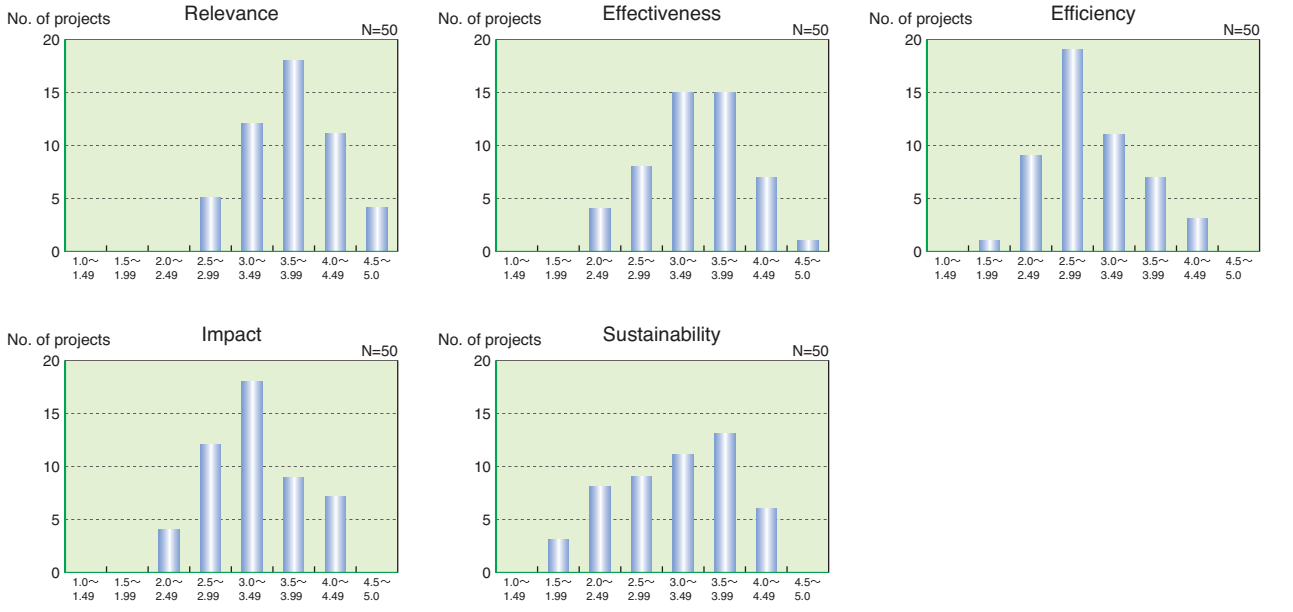
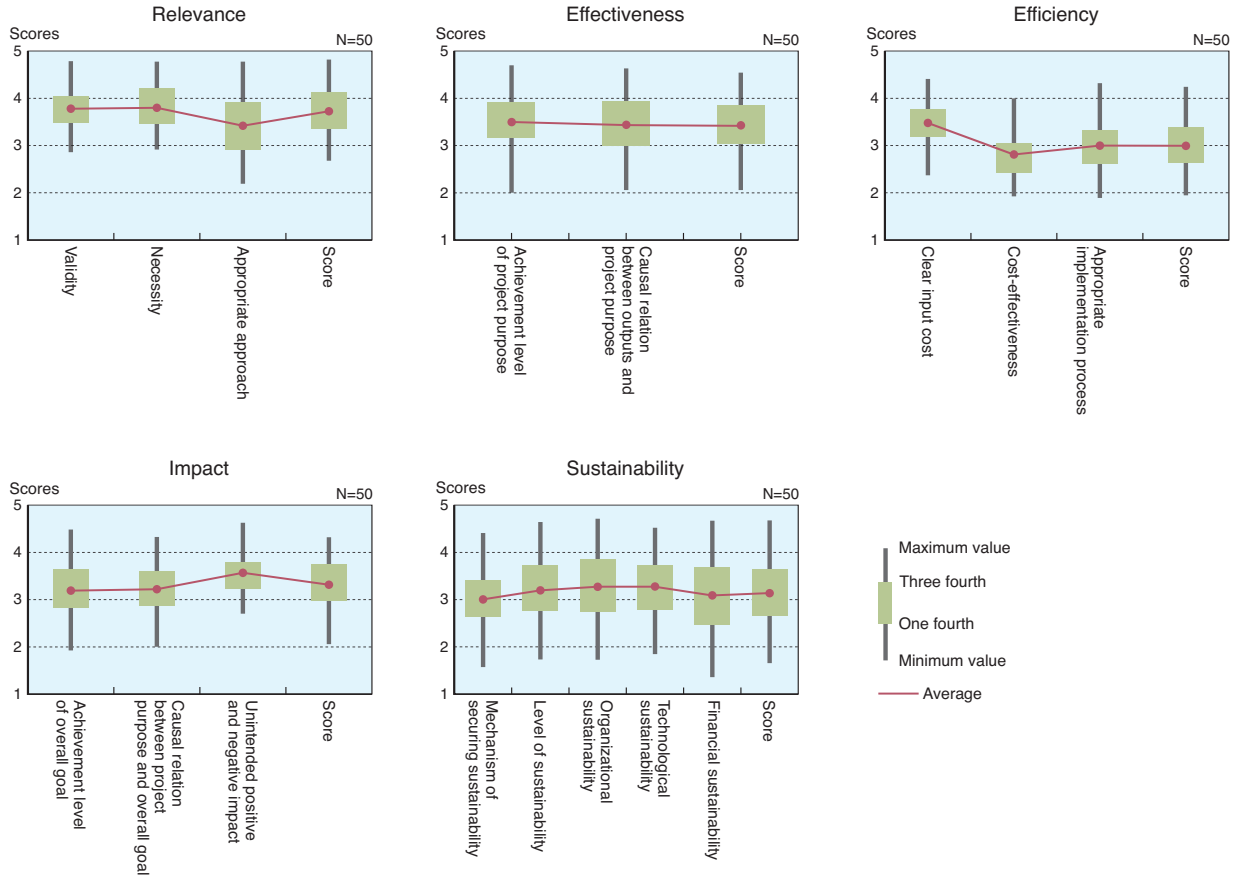


Figure 4-11 Project Evaluation by Secondary Evaluators: Viewpoints (Average Scores and Distributions)



Many projects are of high "validity" and "necessity." Among those projects, highly evaluated projects in terms of "appropriate approach" were those with appropriate project designs, including a design of establishing a regional water supply model with a purpose of building a sustainable waterworks system, and attempting to diversify production activities once operation of the organization ran on track. Conversely, "appropriate approach" was evaluated low for (1) projects that have "necessity" but failed to attract the attention of the government of the partner country, and (2) projects for which the implementing agency in the partner country failed to show interest in cooperation that could have been estimated in advance to some extent in view of the economy, politics, and social circumstances of the partner country.

2) Effectiveness

In the criterion of “effectiveness,” “achievement level of project purpose” and “causal relationships between outputs and project purpose” received a relatively high average score of 3.4 points or higher.

High scores are given to projects that clear the indicators representing achievement level of project purposes, such as using existing technology, establishing new technology that matches the environment, and expanding the activities to profit-seeking enterprises for beneficiaries, thus achieving the project purposes on the grounds of results of output achievement. Conversely, low scores are given to projects implemented as successor projects to those implemented in collaboration with other donors, because of the difficulty of judging whether observed outcomes are due to the foregoing project or the project itself.

3) Efficiency

In the criterion of “efficiency,” the average score for “clear input cost,” a viewpoint to measure if the input cost such as for the purchase of equipment and the dispatch of experts was made clear, is relatively high at 3.5 points. However, the score for “cost-effectiveness” to determine whether efforts were made to achieve more effects with lower costs is low at 2.8 points, which is the lowest rating of all viewpoints under all evaluation criteria.

For “efficiency,” high scores are given to (1) projects that devised a meticulous project plan to raise efficiency, such as shortening the dispatch period of experts, and made a comparision with other donors, and (2) projects that increased cost performance by only dispatching short-term experts, and cut down the costs by meticulous timing of input and effective utilization of the equipment provided in previous projects. In contrast, low scores were given in “efficiency” to (1) projects where a delayed dispatch of Japanese experts resulted in an overall of all activities, (2) projects where the replacement of the mayor or counterparts during the project period caused a delay in establishing the project implementation system, (3) projects where a dispatch of experts was

delayed due to the difficulty in recruiting experts, and failed to smoothly implement the transfer of technology, and (4) projects where the efficiency of project management was hampered due to the frequent replacement of long-term experts, and many research activities were terminated with no prospects for dissemination.

4) Impact

Among the three viewpoints under the criterion of "impact," the average score for "achievement level of overall goal" to determine whether the planned effects were realized or likely to be realized by achieving the project purposes, and "causal relation between project purpose and overall goal" was 3.2 points. On the other hand, the average score for "unintended positive and negative impact" to gauge the emergence of such unintended impact as economic impact on policies, the target society and beneficiaries, and gender equality was highest at 3.6 points, marking a relatively high evaluation when comparing the evaluations of viewpoints in other evaluation items.

High score for "impact" is given to a project, which is highly evaluated with likeliness of achieving its overall goal, dissemination of a sustainable waterworks system by achieving the project purpose of establishing waterworks system, and with many positive impacts it exerts on policies, gender and the environment.

5) Sustainability

The scores of all viewpoints in "sustainability" averaged 3.0 to 3.3, and thus reach the "medium" level or higher. However, the average score for "mechanism of securing sustainability" to evaluate whether a mechanism is incorporated in a project in order to secure sustainability is 3.0, the lowest score among all evaluation viewpoints of sustainability. It is also lower than the scores of viewpoints in other evaluation criteria.

High scores in "sustainability" are given to projects aiming to provide high schools with a new training system based on the dissemination model for training core engineers. This is because such projects were implemented in view of not only the educational sector but also connections with the industrial sector from the beginning. In other words, a mechanism is ready for ensuring sustainability, and high sustainability is expected based on the fact that a teacher training center was constructed to respond to a new system and the Education Ministry allocated a budget necessary for equipment maintenance.

(3) Project Evaluation by Sector

Projects are implemented across a variety of sectors. Those targeted for secondary evaluation in fiscal 2007 can be categorized as: 23 projects in the sector of social development, 10 in agricultural development, nine in forestry and natural environment, four in health and medical care, and four in mining and industrial development. The projects in social development include the development of human resources, regional development, and pollution prevention. Agricultural development includes the promotion of agriculture, rural environment conservation, water management, irrigation, and livestock/veterinary medicine. The projects in forestry and natural environment include forest conservation and management, and utilization of marine resources. The projects in health and medical care include the improvement of regional medical care and measures against AIDS. Those in mining and industrial development include improvements in metal mold technology and productivity.

Although the number of projects evaluated varies from sector to sector, a similar evaluation tendency for the projects is observed. In every sector, the average scores of "efficiency" and "sustainability" are relatively low compared with those of "relevance," "effectiveness," and "impact." In the sectors of agricultural development and mining/industrial development, the average scores for every evaluation criterion are 3.2 or higher, which is a relatively high evaluation compared with other sectors.

The project which obtained the highest score in the sector of agriculture is "Model Planning Project for Water-saving Measures on Large-scale Irrigation Scheme" in China. This project aimed to establish a water-saving irrigation technology that can be spread nationwide through demonstration in the prioritized model irrigation area. This was because, in the midst of a serious water shortage, the use efficiency of agricultural water (accounting for 70% of total water use) had declined due to aging irrigation facilities and inappropriate water management. This project was linked with the project of China's Ministry of Irrigation. Its purposes will be achieved by improving water management technology, developing water-saving technology for application in paddy fields, and developing survey plan methods for the formulation of a water-saving improvement plan. The overall goals are highly likely to be achieved as well, with many positive impacts generated. This project has achieved a weighted overall score of 4.1 and also obtained high scores exceeding 4.0 in the evaluation criteria of "relevance," "effectiveness," "impact" and "sustainability."

In the health and medical care sector, the scores of "relevance" averaged a relatively high 3.7, while those of "effectiveness," "efficiency," "impact" and "sustainability" averaged a low level in the 2-point range. These evaluation criteria are all lower than in other sectors. In "effectiveness," "impact" and "weighted overall score," a statistically significant difference is seen with scores of those in the sector of agricultural development. Only four projects in health and medical care can be evaluated. Among them, the lowest score is given to "Cross Border Initiative Project" in Zambia that aimed to reduce the HIV infection percentage of high-risk groups living along the borders. The scores in "effectiveness," "efficiency," "impact" and "sustainability" are all in the 2-point range. This project is peculiar since it was conducted in collaboration with other donors, and distinguishing JICA's contribution from the overall outcome proved difficult. Thus,

the scores are considerably lower than those of the other three projects. Therefore, we dropped this project, and took average scores of the other three for comparison with other sectors. The results showed somewhat lower scores in health and medical care, but no statistically significant difference.

(4) Project Evaluation by Region

Projects are widely implemented across regions. Twenty-two projects in Asia and Oceania were subject to secondary evaluation in fiscal 2007, along with 14 in Latin America, four in the Middle East, nine in Africa, and one in Europe. Figure 4-12 and Table 4-7 show the evaluation results by region. The Middle East and Europe were merged into one group for the purpose of analysis.

As for the average scores by region, “sustainability” was rated higher in the Middle East and Europe than in Latin America and Africa, with a significant difference statistically. As shown in Figure 4-12, “efficiency” was rated lowest in the Middle East and Europe, Asia and Oceania, and Latin America, while “sustainability” was rated lowest in Africa where the evaluations received low marks for most evaluation criteria compared to those in other regions.

The African region is generally given low scores. The region is home to both high-scoring and low-scoring projects, and shows greater variances in scores among projects than in other regions. The highest-scoring project - "Project on the Safe Water and the Support on Community Activities" in Senegal - is implemented with the purpose of establishing a sustainable waterworks system, through operational guidance given to the water management association and a supporting activity to improve the livelihood of local residents. In the project, Japan reversed the conventional policy centering on hardware in its water supply sector into software for sustainable development, namely maintenance and management, and application and development. The input was effectively applied in the project activities, and the project achieved most of the target values of its indicators during the project period, thereby achieving the project purpose. Moreover, it produced many positive impacts as well. As for each evaluation criterion, the overall score was 4.2, while scores for "relevance," "effectiveness," "efficiency" and "impact" also obtained 4.2 or above. Conversely, four of the nine projects implemented in Africa obtained low overall scores in the 2-point range. The lowest-scoring project - "Project for Strengthening National Bureau of Statistics" in Tanzania is a technical cooperation project for enhancing capacity of the National Bureau of Statistics in the collection, management and provision of statistical data, thus allowing it to provide reliable statistical data at appropriate times, in response to a request for effective implementation of poverty monitoring as part of the state poverty reduction strategy paper. There was a gap between the project outputs and project purpose as stated in the PDM, and it was difficult to achieve the project purpose even upon achieving outputs. It was also hard to recruit and dispatch experts at an appropriate timing, resulting in poor scores for "effectiveness" or "efficiency."

Figure 4-12 Evaluation by Region (Average)

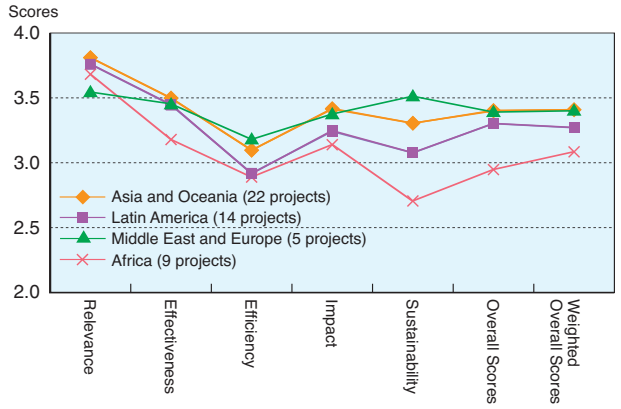


Table 4-7 Evaluation by Region (Average)

	Asia and Oceania	Latin America	Middle East and Europe	Africa	Difference in Average Scores
Relevance	3.82	3.76	3.54	3.68	
Effectiveness	3.49	3.45	3.47	3.17	
Efficiency	3.09	2.91	3.17	2.88	
Impact	3.42	3.24	3.37	3.14	
Sustainability	3.30	3.07	3.51	2.70	*
Overall Scores	3.39	3.30	3.39	2.94	
Weighted Overall Scores	3.40	3.26	3.41	3.08	
Total	22	14	5	9	

* The difference in significance level among regions is 5% on average.

(5) Overall Evaluation of Project by Secondary Evaluators

Fiscal 2007 saw secondary evaluators comprehensively evaluate project achievement by using two methods based on the reports. In one method, the secondary evaluators initially evaluated a project comprehensively regardless of the results of the DAC's five evaluation criteria derived from the report (overall score). In another method, the secondary evaluators weighted and summed up each score of the five evaluation criteria (weighted overall score).

Table 4-8 lists the projects in order from high to low weighted overall scores. The parenthesized numbers in the table denote ranks based on overall score. The projects ranked by using these two methods show that among the top five projects as evaluated in overall scores and the top five projects in weighted overall scores, four projects are taken up in both. For the worst five projects as well, four projects are included in both categories of the worst five projects. As is known from these results, the correlation is high (r = 0.957) between the overall scores and weighted overall scores. When an entire project is evaluated, scoring need not be conducted independently for overall scoring. Instead, it is possible and practical to use the sum of weighted scores of the five evaluation criteria as the overall score for the project.

The projects in each fiscal year were rated with the weighted overall scores. Since the weighted overall score of 3.0 is of the "medium" level, rankings and meanings were given as follows:

The weighted overall scores are:

4.0 or more	A: Excellent project
3.5 to 3.99	B: Good project
2.5 to 3.49	C: Fair project
2.0 to 2.49	D: Partially weak project
Less than 2.0	E: Weak project

The top five and worst five projects were selected from among 50 projects in fiscal 2005 and fiscal 2006 using weighted overall scores, all of which were subject to this fiscal year's secondary evaluation. Figures 4-13 and 4-14 show the scores for those projects in terms of the five evaluation criteria. Table 4-9 compares the averages of weighted overall

Table 4-8 Overall Evaluation of Project by Secondary Evaluators Based on Terminal Evaluation Reports: Weighted Overall Score

Rank	Country	Project Title	Weighted Overall Score
1 (1)	Malaysia	The Project for the Capacity Building of National Institute of Occupational Safety and Health	4.23
2 (9)	Viet Nam	Forest Fire Rehabilitation Project	4.19
3 (2)	Senegal	Project on Safe Water and the Support of Community Activities	4.15
4 (5)	China	The Model Planning Project for Water-saving Measures in Large-scale Irrigation Scheme	4.12
5 (4)	Turkey	Geological Remote Sensing Project	4.12
6 (3)	Thailand	Appropriate Technology for Reduction of Agrochemicals in Northern Thailand	4.09
7 (6)	Turkey	Establishment of Industrial Automation Technologies Departments in Anatolian Technical High Scools	3.93
8 (11)	China	Research and Development Center Project on Sustainable Agricultural Technoloy	3.86
9 (13)	Brazil	Strengthening Agricultural Technical Support System for Small Scale Farmers in Tocantins State Project	3.79
10 (12)	Trinidad and Tobago	The Project for Promotion of Sustainable Marine Fisheries Resource Utilisation	3.74
11 (14)	Niger	School For All	3.72
12 (7)	Brazil	The Project for Forest Conservation and Environmental Education in the Eastern Amazon	3.71
13 (17)	Indonesia	Coal Mining Technology Enhancement Project at Education and Training Unit for Underground Mining	3.64
14 (19)	China	Human Resource Development of Rehabilitation Professionals	3.62
15 (23)	Honduras	Project for the Improvement of Teaching Method in Mathematics	3.62
16 (8)	Thailand	The Project of the Japan-Thailand Technical Cooperation on Animal Disease Control in Thailand and Neighboring Countries	3.62
17 (16)	Viet Nam	Enhancing Capacity of Vietnamese Academy of Science and Technology in Water Environment Protection	3.61
18 (10)	China	The Sino-Japan Friendship Center for Environmental Protection Project (Phase 3)	3.58
19 (15)	Argentina	The Project of Research and Development of Pejerrey Aquaculture and Propagation	3.52
20 (21)	Eritrea	Basic Training for Reintegration of Demobilized Soldiers	3.51
21 (26)	Indonesia	The Empowerment of Water Users Association Project	3.51
22 (22)	Dominican Republic	The Technology Improvement Project for Irrigated Agriculture	3.47
23 (20)	Chile	Strengthening Japan-Chile Partnership Program (JCPP)	3.46
24 (33)	Philippines	Philippine Coast Guard Human Resource Development	3.37
25 (18)	Viet Nam	Japanese Technical Cooperation in the Legal and Judicial Field (Phase 3)	3.28
26 (27)	Indonesia	The Project for Strengthening Decentralized Environmental Management System	3.26
27 (24)	Chile	Rehabilitation for Disabled People Project	3.21
28 (25)	Syria	Project for Capacity Building of Faculty of Veterinary Medicine, Al Baath University	3.20
29 (32)	Panama	Panama Canal Watershed Conservation Project	3.14
30 (36)	Zambia	HIV/AIDS and Tuberculosis Control Project	3.07
31 (34)	Guatemala	Project for Vector Control for Chagas Disease	3.05
32 (30)	Pakistan	Balancing and Modernization of Workshop Facilities at PITAC, Lahore	3.02
33 (29)	Philippines	Gender Responsive Employability (Wage & Self) and Training	3.01
34 (38)	Ethiopia	Capacity Building of the Alemgena Training and Testing Center of the Ethiopian Roads Authority (ERA)	3.00
35 (28)	Micronesia	Extension of the Fisheries Training Project	2.99
36 (43)	Turkey	Technical Development of Sustainable Seed Production for Black Sea Turbot	2.98
37 (35)	Costa Rica	Project on Productivity Improvement for Enterprises	2.97
38 (39)	Thailand	The Project on the Strengthening of Anti-Corruption Capacity	2.95
39 (41)	Cambodia	Technical Service Center for Irrigation System Project	2.94
40 (37)	Bolivia	The Project for Strengthening Regional Health Network of Santa Cruz Prefecture	2.86
41 (42)	Indonesia	Freshwater Aquaculture Development Project	2.83
42 (40)	Romania	The Project on Reduction of Seismic Risk for Buildings and Structures	2.81
43 (44)	Ethiopia	Participatory Forest Management Project in Belete-Gera Regional Forest Priority Area	2.79
44 (31)	Honduras	Project for the Promotion of Self-management Enterprises of Women in Rural Area	2.77
45 (48)	Indonesia	Technical Cooperation for Community Empowerment Program with Civil Society	2.60
46 (46)	Kenya	African Institute for Capacity Development (Phase 2)	2.58
47 (47)	Indonesia	Integrated Sediment-related Disaster Management Project for Volcanic Areas	2.49
48 (45)	Zambia	Cross Border Initiative Project	2.48
49 (50)	Tanzania	Strengthening of National Bureau of Statistics in Data Providing Service	2.41
50 (49)	Mexico	Assistance Plan for Small Producers in El Soconusco Region, the State of Chiapas	2.38

A: Excellent project B: Good project C: Fair project D: Partially weak project

Remarks: The parenthesized numbers in the table denote ranks based on overall score.

Figure 4-13 Top 5 Projects

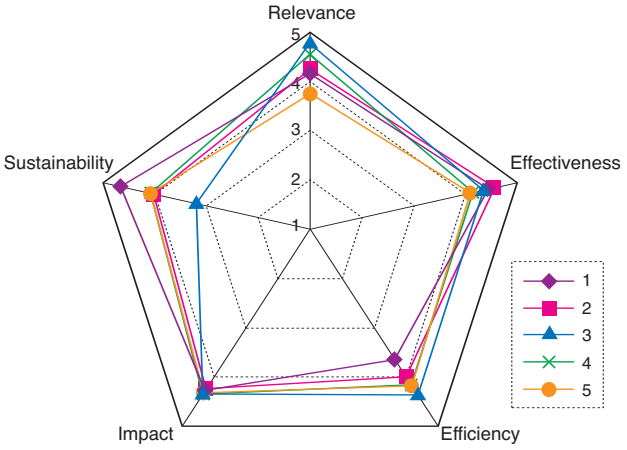


Figure 4-14 Worst 5 Projects

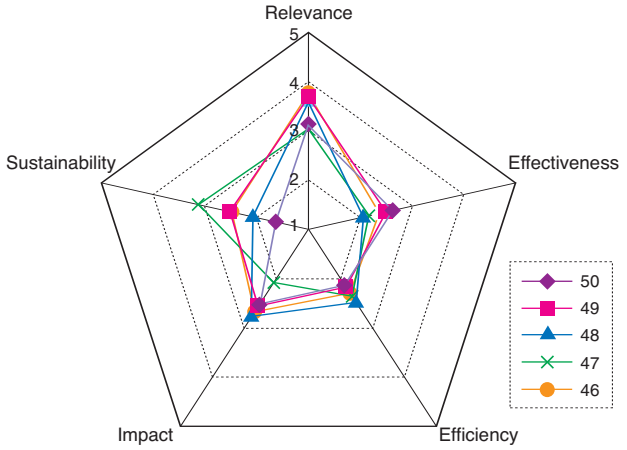


Table 4-9 Differences between Top 5 and Worst 5 (Averages of Weighted Overall Scores)

Evaluation Criteria/Viewpoints		Average Scores: Top 5 Projects	Average Scores: Worst 5 Projects	Difference in Average Scores
Relevance	Validity	4.37	3.36	1.01 **
	Necessity	4.36	3.41	0.95 **
	Appropriate approach	4.31	2.73	1.59 **
	Score	4.37	3.32	1.05 **
Effectiveness	Achievement level of project purpose	4.40	2.36	2.04 **
	Causal relationship between outputs and project purpose	4.38	2.39	1.99 **
	Score	4.32	2.32	2.01 **
Efficiency	Clear input cost	4.13	2.86	1.27 **
	Cost-effectiveness	3.44	2.38	1.05 **
	Appropriateness of Implementation process	3.98	2.12	1.86 **
	Score	3.92	2.23	1.69 **
Impact	Achievement level of overall goal	4.22	2.23	1.99 **
	Causal relationship between project purpose and overall goal	3.84	2.41	1.43 **
	Unexpected positive and negative impact	3.99	3.06	0.93 **
	Score	4.19	2.46	1.74 **
Sustainability	Mechanism of securing sustainability	3.79	2.27	1.51 **
	Level of sustainability	4.03	2.32	1.71 **
	Organizational sustainability	4.15	2.45	1.70 **
	Technological sustainability	4.12	2.54	1.58 **
	Financial sustainability	3.87	2.39	1.48 **
	Score	4.03	2.38	1.65 **
Weighted overall score		4.16	2.47	1.69 **

* The difference in significance level between the average scores of the top and worst five projects is 5%.
** The difference in significance level between the average scores of the top and worst five projects is 1%.

scores for the evaluation criteria and viewpoints for the top five projects and worst five projects.

As shown in Figure 4-13, Figure 4-14, and Table 4-9, the average scores for “relevance,” “effectiveness,” and “impact” of the top project group received high evaluation at 4.2 points or higher. Those for “efficiency” and “sustainability” are 3.9 and 4.0, respectively, and still high but relatively lower than the other three factors. Those for “sustainability” are between 3.2 and 4.7 depending on the projects, showing a variance in ratings. Both evaluation criteria received low evaluation even in the results of secondary evaluation in fiscal 2006.

As for the worst project group, the average score for “efficiency” is 2.2 (lowest among all evaluation criteria), and those for “sustainability” shows a variance in ratings.

Comparing the differences in average scores between the top project group and worst project group, there is a statistically significant difference for each evaluation criterion, with the top project group rated higher than the worst one. Between both groups, the difference in average scores for “relevance” is small, while those for “effectiveness” and “impact” are large.

Next, the differences between the “good” projects and “bad” projects are analyzed based on each evaluation criterion. Comparing the average scores between both groups only reveals a small difference in "relevance" but also shows different magnitudes of difference depending on the viewpoint. The top project group scores higher than 4.3 in every viewpoint and thus achieve high scores. However, the worst project group has an average score of 3.4 for "validity" and "necessity," but shows a relatively low average score of 2.7 for "appropriate approach." Among the viewpoints of "relevance," "appropriate approach" shows a great difference between the two groups. As for "effectiveness," great differences in average score between both groups are shown in the viewpoints of "achievement level of project purpose" and "causal relationship between outputs and project purpose." "Appropriateness of implementation process" in "efficiency" shows a relatively low average score in the low-ranking project group and a great difference from the higher-ranking project group. On the other hand, the average score of "achievement level of overall goal" in "impact" is relatively low in the low-ranking project group, and reveals a great difference from the high-ranking project group. As for "sustainability," the average scores differ greatly in any viewpoints between the two groups. Among them, the average scores of "level of sustainability" and "organizational sustainability" show great differences from the high-ranking project group.

These findings suggest that the appropriate means used in implementing a project, a high achievement level of project purpose, an appropriate implementation process, an assumed impact generated, and high sustainability lead to high project scores.

The highest-ranking project in overall scores is "Project for the Capacity Building of the National Institute of Occupational Safety and Health" in Malaysia, which was also evaluated as the highest in terms of report quality. The scores for evaluation criteria of this project show that "efficiency" scored 3.6, while the scores are 4.3 or higher for "relevance,"

"effectiveness," "impact," and "sustainability." The reasons for these high scores are as follows: All planned activities had been implemented, with outputs achieved. The project purpose of "capacity (technical support, human resource development, collection and dissemination of information) of the National Institute of Occupational Safety and Health is upgraded" had been accomplished, thereby ranking high for "effectiveness." Although external factors are also involved to achieve the overall goal of reducing occupational accidents and deseases in industries, "impact" is great because occupational accidents and diseases were reduced during the project period, more guidelines were issued by the Department of Occupational Safety and Health, and more doctors were participating in the Society of Occupational and Environmental Medicines. The Department of Occupational Safety and Health is expected to continue serving as the central organ, with the National Institute of Occupational Safety and Health continuing its trend of expansion, along with annual increases in project running costs. Moreover, maintaining its technical level and reputation as a research institute is expected to be maintained, and thus "sustainability" is rated high.

The project given the lowest total scores was "Assistance Plan for Small Producers in El Soconusco Region, the State of Chiapas" in Mexico. This project aimed at launching a rural development project under an initiative at both the city and community levels in a model region to support small-scale farmers mainly for women in the state of Chiapas, where poverty indices were the lowest in Mexico. As a rural development project, women groups were engaged in "improved cooking stoves," "sewing," "confectionary," and "flower culture" under the supervision of Japan Overseas Cooperation Volunteers (JOCV). However, the replacement of the city mayor and city officials during the project period caused a delay in establishing the system for project implementation. Moreover, the dispatch of long-term experts for management and the formulation of guidelines for rural development were delayed. There was no clear action plan upon initiating the project.

The scores of the evaluation criteria indicated that "relevance" was 3.3, while scores were in the 2-point range for "effectiveness," "efficiency," and "impact." "Sustainability" obtained the lowest score at 1.7. The reasons for the low scores were as follows: The project achieved the indicators for the project purpose of initiating mini projects in communities other than the pilot one, but one of the chief outputs of the guidelines to improve the management of projects has not been compiled. Consequently, "effectiveness" was slightly low. The factors that impeded project efficiency, such as the delay in dispatching experts, a shortage of JOCV members, and delays in carrying out monitoring, have reduced "efficiency." Unexpected impacts have been made at the community level, such as gender, social equality, and changes in technical aspects. However, such impacts are unlikely to spread to other regions and achieving the overall goal is expected to be difficult. Therefore "impact" is not so high. As for "sustainability," the capacities of the rural development team in the city office were insufficient. Moreover, the external factor of the mayoral election, which takes place every three years, greatly affected the project. Therefore, "sustainability" after completion of the project is therefore expected to be low.

(6) Project Evaluation by Year

Figure 4-15 and Table 4-10 show changes in the average scores for project evaluation by year. The comprehensive evaluation of projects is described with weighted overall scores.

As shown in Figure 4-15, all average scores for all evaluation criteria of projects in fiscal 2004, 2005, and 2006 are higher than those in fiscal 2003. Table 4-10 lists the results of statistical analysis. Comparing the projects in fiscal 2003 to those in fiscal 2004, the latter are rated higher for all evaluation criteria, along with statistically significant differences in

Figure 4-15 Year-to-Year Changes of Project Evaluation by Secondary Evaluators (Average Score)

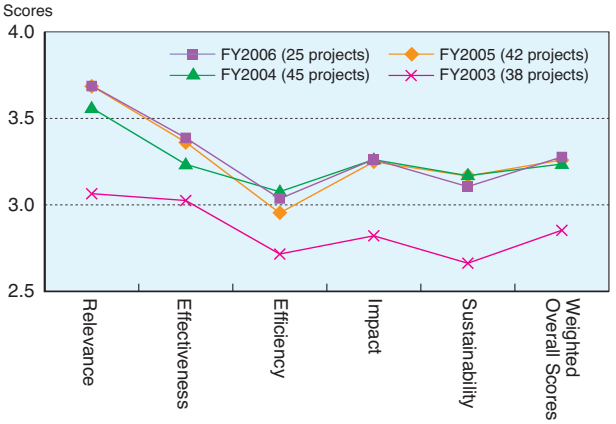


Table 4-10 Year-to-Year Changes of Project Evaluation by Secondary Evaluators (Average Score)

	Average Scores				Difference in Average Scores between Years					
	(1) FY2003	(2) FY2004	(3) FY2005	(4) FY2006	(2)–(1)	(3)–(1)	(4)–(1)	(3)–(2)	(4)–(2)	(4)–(3)
Relevance	3.06	3.56	3.68	3.69	0.50**	0.62**	0.63**	0.13	0.13	0.00
Effectiveness	3.02	3.23	3.36	3.38	0.22	0.35**	0.36**	0.13	0.15	0.02
Efficiency	2.71	3.07	2.95	3.04	0.35**	0.24*	0.32*	–0.11	–0.03	0.08
Impact	2.81	3.25	3.25	3.26	0.44**	0.43**	0.45**	–0.01	0.01	0.01
Sustainability	2.66	3.16	3.16	3.10	0.50**	0.51**	0.44*	0.00	–0.06	–0.06
Weighted Overall Scores	2.85	3.23	3.26	3.27	0.38**	0.41**	0.42**	0.03	0.04	0.01

* The difference in significance level between the scores in fiscal years is 5% on average.
** The difference in significance level between the scores in fiscal years is 1% on average.

"relevance," "efficiency," "impact" and "sustainability" between these projects. Comparing the projects in 2003 to those in 2005 and 2006, the projects in 2003 are rated lower for all evaluation criteria, and a statistically significant difference can be observed. Moreover, from the viewpoints of weighted overall scores that show a comprehensive evaluation of projects, statistically significant differences can also be observed between 2003 and 2004, 2003 and 2005, and 2003 and 2006, where the scores in 2003 rated lower. There are no significant differences between 2004, 2005 and 2006.

Weighted overall scores were used to rate the projects in each fiscal year. The highest rating is A, followed by B, C, D, and E in descending order. Weighted overall scores higher than 4.0 are rated as A, 3.5 to 3.99 as B, 2.5 to 3.49 as C, 2.0 to 2.49 as D, and less than 2.0 as E. Figure 4-16 shows the distribution by project rating. As is evident from this figure, fiscal 2003 shows projects of E, while fiscal 2004, 2005, and 2006 show no such projects. Projects of D account for 23.7% of all the projects in fiscal 2003, while the percentage drops annually from fiscal 2004 to 2005 to 2006. Comparing fiscal 2003 to 2004 shows that fiscal 2003 has 2.6% of projects of B, while fiscal 2004 witnessed a rise to 29.5%, even with projects of A. Comparing fiscal 2004 and 2005 shows slightly fewer projects of D in fiscal 2005. Projects of A and B have similar percentages in both fiscal years, but projects of A have risen from 2.3% in fiscal 2004 to 9.5% in fiscal 2005. Comparing fiscal 2005 to 2006 shows a slight decline in projects of D in fiscal 2006, with projects of A accounting for a similar percentage of magnitude. As for projects of B, fiscal 2005 showed a percentage of 23.8% and fiscal 2006 showed 36.0%, thereby indicating more projects with high rating in fiscal 2006. Comparing the averages of weighted overall scores reveals unclear changes in evaluations conducted in fiscal 2004, 2005, and 2006. Conversely, a look at the distribution of project rating reveals that projects of C and higher increased annually in percentage. It can be concluded that projects are given higher scores as the years go by, though these scores are rated from the reports by secondary evaluators.

Figure 4-16 Year-to-Year Changes of Project Evaluation by Secondary Evaluators (Distribution)

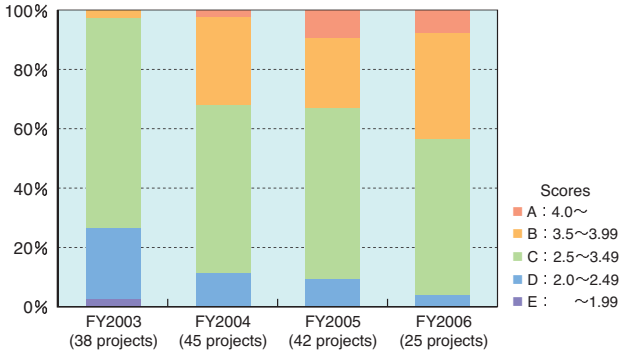
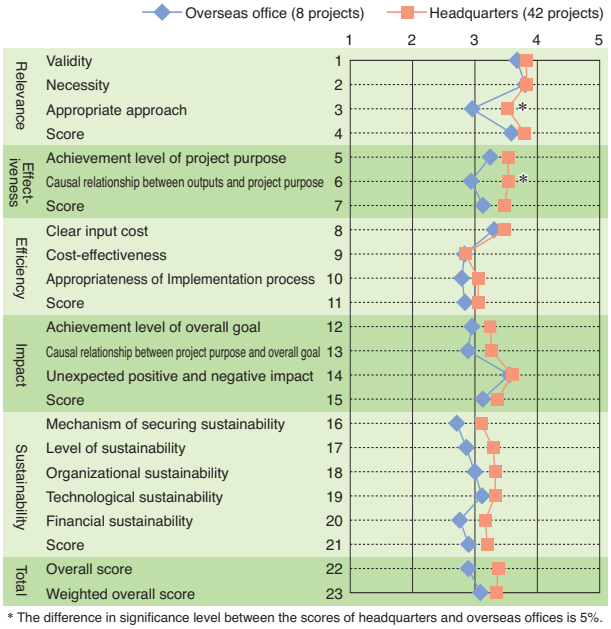


Figure 4-17 Project Evaluations by JICA Headquarters and Overseas Offices



* The difference in significance level between the scores of headquarters and overseas offices is 5%.

(7) Evaluations by JICA Headquarters and Overseas Offices, and Introduction of Ex-ante Evaluation

In line with the trend toward decentralization, JICA adopted "direct project management by overseas offices" in October 2004. Under this system, the overseas offices are authorized to conduct a series of operations from project formulation to project evaluation. Since April 2005, the number of supervising offices has been expanded to total 30 offices. Of the projects subjected to secondary evaluation, two in fiscal 2005 and six in fiscal 2006 were subjected to terminal evaluation at overseas offices. Figure 4-17 shows scores classified by projects subjected to terminal evaluation by JICA headquarters and overseas offices.

There were as few as eight projects subjected to terminal evaluation by overseas offices, and thus posed a constraint. However, as is evident from this chart, projects by headquarters tend to generally rate somewhat higher than projects by overseas offices in terms of project evaluations by secondary evaluators based on the reports. For "appropriate approach" concerning "relevance," and for "causal relationship between outputs and project purpose" concerning "effectiveness," the scores of projects by headquarters are significantly higher statistically than projects by overseas offices. As stated in "quality of primary evaluation," the quality of terminal evaluation shows differences in the viewpoints of "evaluability of project plan," "logic of project design," and "project monitoring," and the terminal evaluation by headquarters was higher in quality. This leads us to the following conclusion: projects by headquarters present their project purposes clearly and make appropriate plans toward achieving them when the projects are launched, thus leading to the high scores of the projects. Conversely, two of the worst five projects as scored by secondary evaluators were those subjected to project formulation and terminal evaluation by overseas offices. Achievement of project purposes in these projects had been affected

by such peculiar factors as follows: The project was conducted in collaboration with other donors, and it was difficult to distinguish JICA's contribution from the overall. The project was greatly affected by external factors. However, these low scores of two projects presumably led to the low scores of all projects by the overseas offices.

With regard to ex-ante evaluation, 24 projects were subjected to ex-ante evaluation and 26 were not among the projects subjected to secondary evaluation in fiscal 2007. In the evaluation results for projects by secondary evaluators, there were no evaluation criteria that show statistically significant differences between the projects subjected to ex-ante evaluation and those not subjected to it.

(8) Summary of Project Evaluation by Secondary Evaluators Based on the Reports

In general, “relevance” of the target projects was high, and other factors of “effectiveness,” “impact,” and “sustainability” were evaluated as “medium” or higher. “Efficiency” also achieved a “medium” level, but its evaluation was relatively low.

An analysis of the DAC's five evaluation criteria by viewpoint reveals some differences in the viewpoints even in the same evaluation criterion. As for "relevance," high scores were given to "validity of project implementation" that evaluates conformity with Japan's aid policy, JICA Country Programs, and aid policy of the partner country, or the high priority of project implementation. "Necessity of the project" that evaluates the target group, target region, and consistency of social needs was high. However, low scores were given to the viewpoint of "appropriate approach," such as the appropriateness of the project as an effective approach toward solving the development issues, the appropriateness of selecting a target region and target group, and the advantage of Japan's technology. As for "efficiency," relatively high scores are given to the viewpoint of "clear input cost" that determines whether the input cost such as for the procurement of equipment and dispatch of experts is clearly grasped. However, low scores are given to the viewpoint of "cost-effectiveness" that determines whether efforts were made to minimize costs, whether there was an alternative means of achieving the purpose at lower cost, whether a higher achievement could be realized with the same cost, and whether cost-effectiveness of a particular project was higher than that of similar projects. Also, low scores are given to the viewpoint of "appropriate implementation process" that concerns the appropriateness of the timing, scale, and quality of input. In other words, the scores failed to reach the "medium" level. For "sustainability" as well, the "medium" level is reached in "mechanism of securing sustainability" that determines whether efforts and mechanisms for securing sustainability were considered in the project, and "financial sustainability" that determines whether adequate measures were taken to secure a sufficient budget. However, the scores given for both are relatively low.

A chronological comparison of the evaluation results of projects subjected to terminal evaluation from fiscal 2003 to 2006 reveals statistically significant differences in average scores between fiscal 2003, and 2004, 2005, 2006. The projects in fiscal 2004, 2005, and 2006 are higher in quality than those in fiscal 2003. However, no differences were found among the projects in fiscal 2004, 2005, and 2006. In the analysis of fiscal 2007, each project was given a sum of weighted scores of DAC's five evaluation criteria as a weighted overall score, and rated in five grades from "A: Excellent project" to "E: Weak Project." As viewed from the distribution of project ratings, the percentage of projects of C and higher tend to rise as the years go by. As evaluated by secondary evaluators based on the reports, it can be concluded that projects are rated higher as the years go by.

Comparing projects generally rated higher by secondary evaluators to projects rated rather low in general, great differences were indicated in "appropriate approach" regarding "relevance," "achievement level of project purpose" and "causal relationship between outputs and project purpose" regarding "effectiveness," "appropriate implementation process" regarding "efficiency," "achievement level of overall goal" regarding "impact," and "level of sustainability" and "organizational sustainability" regarding "sustainability." In project management, it is necessary to consider these points.

Comparing projects by headquarters with projects by overseas offices, the scores of projects supervised by headquarters generally tend to be slightly higher than those by overseas offices though the scores were given by secondary evaluators based on the reports. For "appropriate approach" concerning "relevance" and for "causal relationship between outputs and project purpose" concerning "effectiveness," projects by headquarters are significantly higher statistically. Concerning the viewpoints of "evaluability of project plan," "logic of project design," and "project monitoring" in connection with the quality of terminal evaluation, terminal evaluation by headquarters is higher in quality. This leads us to conclude the following: Projects by headquarters present their project purposes clearly and formulate appropriate plans toward achieving their project purposes when the projects are launched, thus leading to high scores of the projects. This suggests that it is important to prepare a meticulous plan toward achieving the project purpose at the stage of formulating an implementation plan for the project. In overseas offices, project outcome can presumably be higher when designing an appropriate PDM and implementing evaluation according to JICA's guidelines.

(9) For the Betterment of Quality of Projects

The points for improving the quality of projects including those already described in the secondary evaluation in fiscal 2005 and 2006 can be summarized as follows:

1) Appropriate Approach for Project Implementation

In project implementation, relevance at the initial stage will affect the subsequent implementation process and production of outputs. The scores of "appropriate approach" in "relevance" evaluated by secondary evaluators were lower than those of "validity" and "necessity." It is important for

better project implementation to carefully examine the situation surrounding the project such as external factors, and select the appropriate means and methods. The appropriate means and methods will presumably ensure the achievement of project purpose, generation of impact, and increase in sustainability, and thus enhance project quality.

2) Clear Causal Relationship between the Project Purpose and Overall Goal

The outcome defined as the overall goal emerges when the project purpose is achieved. However, some projects had vague relationships between project purpose and overall goal, or a gap between the overall goal and project purpose. In order to increase achievement of the overall goal, it is necessary to clarify the causal relationships among activities, project purpose, and overall goal.

3) Setting Indicators and Target Values

Some reports may face difficulty in evaluating the level of achievement; for example, indicators to measure the level of achievement of the project activities and purpose were not set, or no clear target values were provided even if the indicators were set. In designing a PDM, it is important to fully examine whether the indicators respond accurately to the activities and purpose. It is also vital to set a target value in order to secure the objectivity of evaluation results.

4) Guidance to overseas offices on how to implement a project

Although evaluated by secondary evaluators through terminal evaluation, the projects by overseas offices tended to show generally lower scores than projects by JICA headquarters. Particularly large differences are noted in scores for "appropriate approach" and "causal relationship between outputs and project purpose" between projects by overseas offices and projects by headquarters. Projects by headquarters present their project purposes clearly and make appropriate plans toward achieving those project purposes, which has presumably led to high scores of the projects themselves. To cope with more projects being implemented by overseas offices, headquarters needs to take measures to improve the quality of projects implemented by overseas offices, such as providing guidance to the overseas offices on formulating a PDM and utilizing evaluation methodology. The overseas offices, in turn, must prepare an appropriate PDM in line with JICA guidelines.

Appendix 1 List of Projects Subject to Secondary Evaluation in Fiscal 2007

Fiscal 2006 (New Targets): 25 projects	
Bolivia	The Project for Strengthening Regional Health Network of Santa Cruz Prefecture
Brazil	The Project for Forest Conservation and Environmental Education in the Eastern Amazon
China	Human Resource Development of Rehabilitation Professionals
China	Research and Development Center Project on Sustainable Agricultural Technology
Eritrea	Basic Training for Reintegration of Demobilized Soldiers
Ethiopia	Participatory Forest Management Project in Belete-Gera Regional Forest Priority Area
Honduras	Project for the Promotion of Self-management Enterprises of Women in Rural Area
Indonesia	Technical Cooperation for Community Empowerment Program with Civil Society
Indonesia	The Empowerment of Water Users Association Project
Kenya	African Institute for Capacity Development (Phase 2)
Niger	School For All
Pakistan	Balancing and Modernization of Workshop Facilities at PITAC, Lahore
Philippines	Gender Responsive Employability (Wage & Self) and Training
Philippines	Philippine Coast Guard Human Resource Development
Romania	The Project on Reduction of Seismic Risk for Buildings and Structures
Syria	Project for Capacity Building of Faculty of Veterinary Medicine, Al Baath University
Tanzania	Strengthening of National Bureau of Statistics in Data Providing Service
Thailand	The Project of the Japan-Thailand Technical Cooperation on Animal Disease Control in Thailand and Neighboring Countries
Thailand	Appropriate Technology for Reduction of Agrochemical in Northern Thailand
Thailand	The Project on the Strengthening of Anti-Corruption Capacity
Trinidad and Tobago	The Project for Promotion of Sustainable Marine Fisheries Resource Utilisation
Turkey	Technical Development of Sustainable Seed Production for Black Sea Turbot
Viet Nam	Enhancing Capacity of Vietnamese Academy of Science and Technology in Water Environment Protection
Viet Nam	Japanese Technical Cooperation in the Legal and Judicial Field (Phase 3)
Viet Nam	Forest Fire Rehabilitation Project
Fiscal 2005 (New Targets): 25 projects	
Argentina	The Project of Research and Development of Pejerrey Aquaculture and Propagation
Brazil	Strengthening the Agricultural Technical Support System to Small Scale Farmers in Tocantins State
Cambodia	The Project for Technical Service Center for Irrigation System
Chile	Enforcement of Japan Chile Partnership Programme (JCPP)
Chile	Rehabilitation for Disabled People Project
China	The Model Planning Project for Water-saving Measures on Large-scale Irrigation Scheme
China	The Sino-Japan Friendship Center for Environmental Protection Project (Phase 3)
Costa Rica	Project on Productivity Improvement for Enterprises
Dominican Republic	The Technology Improvement Project for Irrigated Agriculture
Ethiopia	Project for Capacity Building of ERA Training and Testing Center Alemgena
Guatemala	Project for Vector Control for Chagas Disease
Honduras	The Improvement of Teaching Method in Mathematics
Indonesia	Coal Mining Enhancement Project at Ombilin Mines Training College
Indonesia	Freshwater Aquaculture Development Project
Indonesia	Integrated Sediment-related Disaster Management Project for Volcanic Areas
Indonesia	The Project for Strengthening Decentralized Environmental Management System
Malaysia	The Project for the Capacity Building of National Institute of Occupational Safety and Health in the Field of Occupational Safety and Health
Mexico	Project on the Assistance Plan for Small Producers in El Soconusco Region
Micronesia	The Fisheries Training Project (Extended)
Panama	Panama Canal Watershed Conservation Project
Senegal	Project on Safe Water and Support of Community Activities
Turkey	Establishment of Industrial Automation Technologies Departments in Anatolian Technical High Schools
Turkey	Geologic Remote Sensing Project
Zambia	The Strengthening of Laboratory Systems for HIV/AIDS and TB Control Project
Zambia	Cross Border Initiative Project
Both Years as Seam Allowance (2006 and 2005): 17 projects	
Argentina	The Project on Establishment of Control Capacity for Industrial Wastewater and Waste
Brazil	The Technological Development Project for Sustainable Agriculture in Eastern Amazonia
Brazil	Technology Development for Revegetation and Utilization of Degraded Areas in the Semi-arid Region of the Northeastern Brazil
Chile	The Project on Conservation of the Environment and Rural Development with Farmers' Participation for the Mediterranean Dryland Zone of Chile
China	China-Japan Friendship Project on the National Center for Safety Evaluation of Drugs
China	Research of Performance Assessment and Product Certification for Residential Building
Ghana	Improvement of Educational Achievement in Science, Technology and Mathematics (STM) in Basic Education
Indonesia	The Forest Fire Prevention Management Project (Phase 2)
Indonesia	Establishment and Capacity Building of Regional Export Training and Promotion Centers
Kenya	Promotion of Sustainable Community Based Small-holder Irrigation
Laos	Development of Bases to Autonomously Carry out Reading Promotion Project
Thailand	Development of Vocational Opportunities and Creative Activities for People with Disabilities and Commercializing Hill-tribes Peoples' Crafts
Thailand	Project on Local Management Cooperation
Thailand	The Project on the Practical Energy Management Training Center
Thailand	The Reforestation and Extension Project in the Northeast of Thailand (Phase 2)
Philippines	Promotion of the Ship Inspection System and Technique
Viet Nam	Project on the Improvement of Higher Maritime Education

Appendix 2 Secondary Evaluation Check Sheet (Fiscal 2007)

Evaluator		Date
Project title		

Rating criteria

1) Rate viewpoints and criteria in green cells and orange cells based on a scale of 1 to 5. [I – III]

Criteria
- 5: Sufficient/high
- 4: Fairly sufficient/high
- 3: Average
- 2: Slightly insufficient/low
- 1: Insufficient/low

2) Rate familiarity in green cells choosing from the dropdown list.

3) Write down highlights and notable points (including good practices) in the space for comment. [I – IV]

I. Preconditions for appropriate evaluation (evaluability)

	[Appropriateness of Project Plan (ex-ante evaluation or PDM)] Whether project plan (ex-ante evaluation/PDM) is designed properly?	
Viewpoint	Whether the revised PDM used for the evaluation is a better tool for evaluation than the original. Whether the project purpose in the revised PDM is not set lower than the original.	
	[Target Group]	
Viewpoint	The target group, beneficiaries of the project, is clearly and appropriately set.	
	[Verifiability of Project Purpose] Whether project output and purpose are set properly in the project plan so as to measure the achievement?	
Viewpoint	The indicators and specific target values (parameter) are clearly defined for each output and project purpose.	
	[Verifiability of Overall Goal] Whether the overall goal is set properly in the project plan so as to measure the achievement?	
Viewpoint	The indicators and specific target values (parameter) are clearly defined for overall goal.	
	[Logic of Project Design] Whether "activity," "output," "project purpose," and "overall goal" are relevant logically?	
Viewpoint	The PDM for the evaluation sets a clear and realistic hypothesis and logical flow considering important assumptions.	
	[Verifiability of Performance and Implementation Process] Whether project monitoring is conducted and documentation is properly conducted?	
Viewpoint	Monitoring of outputs, activities, and inputs was regularly conducted, and the information including statistical data is accumulated during project implementation.	
Rating	Comment	

II. Secondary Evaluation for Each Criterion

1. Evaluation Framework		
	[Evaluation Team Composition (Neutrality/Fairness/Expertise)] Whether evaluation team is organized considering neutrality, fairness, and expertise on the premise of the internal evaluation by JICA.	
Viewpoint	Whether it is clearly mentioned in the report about the concerns for the neutrality/fairness/expertise of manpower resource on the premise of the internal evaluation by JICA. Based on the descriptions about the current positions of evaluators and the relationship with the projects and any other information on the evaluation report, judge whether fairness and neutrality are not corrupted, or whether the quality of evaluation is not harmed due to lack of expertise or imbalance of evaluation team composition.	
	[Level of Counterpart Participation] Whether the participation of counterpart is sufficient (as an evaluator)	
Viewpoint	Whether there is a concrete description in the report about the sufficient participation of counterpart to evaluation (= engagement with the understanding evaluation method in the series of evaluation activities by project stakeholders or other third company in the counterpart country).	
Rating	Comment	
2. Data Collection *1		
	[Evaluation Questions] Whether evaluation questions are set properly.	
Viewpoint	Evaluation questions are set in line with evaluation purposes and set properly in the evaluation grid. General questions regarding evaluation criteria are narrowed down to more specific sub-questions to identify necessary information/data to be collected.	
	[Appropriateness of Data Collection Methods and Data Sources] Whether methods and resources for data collection are appropriate.	
Viewpoint	Several different data collection methods are used to increase accuracy and reliability of the data/information obtained. The data/information is obtained from a broad range of stakeholders, including the end beneficiary groups.	
	[Clarity of Data/Information Sources] Whether the data/information sources are clearly referred.	
Viewpoint	The sources of the data/information are adequately explained in the evaluation report in the forms of references and the lists of interviewees.	
	[Sufficiency of Data/Information Obtained] Whether information is sufficiently collected.	
Viewpoint	Data collection is conducted based on the evaluation grid, and the data/information is sufficient to answer the evaluation questions, and additional information/data is gathered for unexpected and newly confronted questions during the evaluation process.	
Rating	Comment	

3. Analysis	
3.1 Assessment of Performance and Verification	
	[Comprehension and Verification of Project Performance (Outputs)] Whether outputs are comprehended and verified sufficiently.
Viewpoint	Achievement level of outputs is measured with the target values set by the indicators.
	[Comprehension and Verification of Project Performance (Project Purpose)] Whether the achievement of project purpose is comprehended and verified.
Viewpoint	Achievement or expected level of project purpose is measured with the target values set by the indicators.
	[Comprehension and Verification of Project Performance (Overall Goal)] Whether the achievement of overall goal is comprehended and verified.
Viewpoint	Achievement or expected level of overall goal is measured with the target values set by the indicators.
	[Comprehension and Verification of Project Performance (Project Implementation Process)] Whether the implementation process is comprehended and verified.
Viewpoint	The project implementation process is thoroughly examined, through which impeding and/or promoting factors to achievement of outputs, project purpose, and overall goal are identified.
	[Comprehension and Verification of Project Performance (Qualitative Causal Relationship—Logic of Project Design)]
Viewpoint	The logic of project design is thoroughly verified, through which impeding and/or promoting factors to achievement of outputs, project purpose, and overall goal are identified*2.
	[Comprehension and Verification of Project Performance (Quantitative Causal Relationship—Before and After)]
Viewpoint	The causal relationships are thoroughly examined to verify that effects for the beneficiaries have resulted from the project interventions*3.
Rating	Comment
3.2 Analysis Method	
	[Objective Analysis] Whether objective analysis is conducted based on data.
Viewpoint	The data is objectively analyzed from the facts based on a series of scientific discussions, and an effort is made to quantify the data where feasible.
	[Holistic Analysis] Whether holistic analysis is conducted.
Viewpoint	The data interpretation is drawn by examination and analysis of various methods and aspects.
	[Analysis of Promoting and Impeding Factors] Whether the analysis of promoting and impeding factors is conducted.
Viewpoint	Factors that promote and impede effects are adequately sufficiently analyzed in light of the project logic (cause-effect) and the project implementation process (such as project management).
Rating	Comment
4. Evaluation by Five Criteria*4	
	[Five Evaluation Criteria (Relevance)] Whether the evaluation on relevance is sufficient.
Viewpoint	Perspectives for evaluation of "Relevance" (validity and necessity of a project in light of needs of beneficiaries, project implementation as an appropriate approach to problem solving, consistency of policies, etc.) are sufficiently covered.
	[Five Evaluation Criteria (Effectiveness)] Whether the evaluation on effectiveness is sufficient.
Viewpoint	Perspectives for evaluation of "Effectiveness" (achievement level of project purpose, causal relationships between outputs and project purpose, etc.) are sufficiently covered.
	[Five Evaluation Criteria (Efficiency)] Whether the evaluation on efficiency is sufficient.
Viewpoint	Perspectives for evaluation of "Efficiency" (comparison with other similar projects through cost analysis, cost-effectiveness analysis, etc.) are sufficiently covered.
	[Five Evaluation Criteria (Impact)] Whether the evaluation on impact is sufficient.
Viewpoint	Perspectives for evaluation of "Impact" (achievement level of overall goal, causal relationships between project purpose and overall goal) are sufficiently covered.
	[Five Evaluation Criteria (Sustainability)] Whether the evaluation on sustainability is sufficient.
Viewpoint	Perspective for evaluation of "Sustainability" (mechanism for securing sustainability and outcomes to be produced in terms of policies and systems, organizational and financial aspects, technical aspects, socio-culture, and environment) are sufficiently covered.
	[Total Evaluation (Conclusion)] Whether conclusion is drawn properly.
Viewpoint	The conclusion is drawn based on holistic viewpoints that are in turn based on the five evaluation criteria
Rating	Comment

5. Recommendations/Lessons Learned*5	
5.1 Recommendations	
	[Recommendations (Sufficiency of Recommendations)] Whether recommendations are drawn sufficiently.
Viewpoint	The recommendations fully consider the impeding/promoting factors identified during the evaluation process.
	[Recommendations (Relevance and Credibility of Recommendations)] Whether recommendations are drawn from the evaluation results and include persuasive contents.
Viewpoint	The recommendations are based on the information obtained through the process of data analysis and interpretation. As a result, the recommendations are objective and convincing.
	[Usability of Recommendations] Whether recommendations are presented well enough to be applicable for future activities.
Viewpoint	The recommendations are practical and useful for feedback and follow-ups, with a specific time frame as well as target of responsibility.
Rating	Comment
5.2 Lessons Learned	
	[Lessons (Sufficiency of Lessons Learned)] Whether lessons are fully drawn.
Viewpoint	The lessons learned fully consider the impeding/promoting factors identified during the evaluation process.
	[Lessons (Relevance and Credibility of Lessons Learned)] Whether lessons are drown from the evaluation result and include persuasive contents.
Viewpoint	The lessons learned are based on the information obtained through the process of data analysis and interpretation. As a result, the lessons learned are objective and convincing.
	[Usability of Lessons Learned] Whether lessons are presented well enough to be applicable for future activities.
Viewpoint	The lessons are generalized and conceptualized so that they are widely applicable in the future.
Rating	Comment
6. Reporting	
	[Presentation (Conciseness, Clarity, Clearness)] Whether the report is presented in a concise and clear manner so that the readers comprehend easily.
Viewpoint	The evaluation report is simple and clear, and understandable to readers—in light of the structure, font, terminology, and data presentation. The PDM is stated in the beginning of the body text. Logical structure and major points are clearly described in an easily understandable manner.
	[Utilization of Tables and Figures] Whether the intentions are presented with tables and figures.
Viewpoint	Tables and figures are effectively utilized to present statistics and analysis results visually.
	[Presentation of Primary Data] Whether the contents and results of interviews/questionnaires are stated
Viewpoint	Sufficient primary data such as those on targets and results of interviews and questionnaires or sources are presented properly in the report.
Rating	Comment

III. Evaluation of the Project Based on the Report

Fill in comments if there are any important assumptions that might affect the following Five Evaluation Criteria.

1. Relevance (Validity and Necessity for Project Implementation)	
	[Validity] Whether there is validity of project implementation.
Viewpoint	The project is consistent with Japan's aid policies, JICA Country Program, and development policies of the partner country. Its implementation in ODA is relevant. The priority of project implementation is high as cooperation in the partner country and target sector.
	[Necessity] Whether there is necessity of project implementation.
Viewpoint	The project is in line with the needs of the target group, area, and society. Those needs are still present and logically understood including priority.
	[Appropriate Approach] Whether project design is appropriate.
Viewpoint	The approach is appropriate to solve the preset development issue (overall goal). The selection of target area and group is appropriate. Japanese technology is superior. To achieve higher level of outcomes, partnership with other donors and the related projects in the partner county is planned and implemented.
Rating	Comment
2. Effectiveness (Achievement of Project Purpose)	
	[Achievement Level of Project Purpose] Whether project purpose is achieved.
Viewpoint	Project purpose has been (is going to be) achieved.
	[Causal Relationships between Outputs and Project Purpose] Whether cause-effect relationship is strong enough.
Viewpoint	Project purpose has been (is going to be) achieved as a result of outputs. Important assumptions which might affect the achievement of outputs and project purpose were properly identified. There were special factors which impeded or promoted effectiveness.
Rating	Comment

3. Efficiency (Efficiency of Project)		
	[Clarity of Input Cost] Whether input cost is comprehended clearly.	
Viewpoint	Unit costs for purchasing equipment and dispatching experts are clearly presented.	
	[Cost-effectiveness] Whether utmost efforts are made for cost-effectiveness	
Viewpoint	Efforts to cut down on costs were made (using local resources).There was no alternative means that could have led to the same achievements at lower costs. It was impossible to produce greater achievements at the same costs. Compared to similar projects of other donors and the partner country, the cost-effectiveness was high.	
	[Appropriate Implementation Process] Whether the implementation process is appropriate.	
Viewpoint	The inputs were made in a timely manner with appropriate scale and quality.	
Rating	Comment	
4. Impact (Expected, Unexpected Effect by the Achievement of Project Purpose)		
	[Achievement Level of Overall Goal] Whether planned effect is attained due to the achievement of project purpose.	
Viewpoint	Effects planned in the project (overall goal) have been achieved (or are likely to be achieved based on clear grounds) as a result of achievement of project purpose. Problem-solving for the target project has progressed.	
	[Causal Relationships Regarding Impact] Whether there are causal relationships between the project purpose attained and expected effect.	
Viewpoint	Impact was generated as a result of achievement of project purpose. There were special factors that promoted or impeded planned effects including important assumptions.	
	[Unexpected Positive and Negative Impact] Whether unexpected positive and negative impacts affect.	
Viewpoint	There are political impacts and economical impacts on the target society, inside the implementing agency, and on the beneficiary. Other impacts on organization, development of related regulation and laws, gender equality, human rights, disparity between rich and poor, peace and war, and environmental protection are present. There are special factors that brought unexpected positive and negative impacts. When there are many unexpected positive impacts, rate 5 and when there is a few, rate 4; when there are many unexpected negative impacts, rate 1, and when there is a few, rate 2; when there are no unexpected impacts, rate 3.	
Rating	Comment	
5. Sustainability (Sustainability after Completion of JICA's Technical Cooperation)		
	[Mechanism of Securing Sustainability] Whether mechanism for sustainability are institutionalized through project implementation.	
Viewpoint	Mechanisms and devices for securing sustainability (management capacity of the implementing agency, policy support from the supervising agency, demand for activities of the implementing agency, securing financial basis) were considered in the project.	
	[Level of Sustainability] Whether the effects would last after the completion of aid.	
Viewpoint	Effects aimed for in the project (project purpose and overall goal) are (will be) sustained after the termination of cooperation.	
	[Organizational Sustainability] Whether there is sufficient capability of organization to secure sustainability.	
Viewpoint	The positioning of activities in the policies and organization of the implementing agency is stable enough to conduct activities that will continue effects after the termination of cooperation.	
	[Technological Sustainability] Whether there are sufficient skills and techniques to secure sustainability.	
Viewpoint	Technology and capacity acquired in the project are maintained and expanded. Equipment is properly maintained and managed.	
	[Financial Sustainability] Whether there is sufficient finance to secure sustainability.	
Viewpoint	Budget including operating expenses is secured. Measures for securing budget are sufficient.	
Rating	Comment	
6. Overall Rating of above-described Five Evaluation Criteria (Feel free to interpret the weighting of each evaluation criterion).		
Rating	Comment	

IV. Familiarity toward the Concerned Project

	Prior Information about the Project				
Viewpoint	1. None	2. Little	3. Some	4. Much	5. Substantial
	Familiarity with Region				
Viewpoint	1. None	2. Little	3. Some	4. Much	5. Substantial
	Familiarity with Expertise				
Viewpoint	1. None	2. Little	3. Some	4. Much	5. Substantial

V. Overall Comment

Notes:

*1 Major data collection methods
1. Literature review
2. Direct observation
3. Questionnaire survey
4. Interview survey
5. Focus group discussion

*2: Qualitative approach to analyze causal relationships
1. Construct information on implementation process from inputs through activities to outputs, and from outputs to objectives
2. Attempt to explain the logical relationship between project implementation and effects
3. Analyze the process to transfer and disseminate technologies through activities
4. Clarify the relationship between project implementation and effects by conducting detailed and in-depth survey of a small target region or small target group (e.g. case study)

*3: Quantitative approach to analyze causal relationships
1. See changes of the target society/ beneficiary after the project
2. Compare the target society/ beneficiary with another society/ beneficiary without the project
3. Combination of 1 and 2 (experimental design method)
4. Combination of 1 and 2 (quasi- experimental design method)

*4: Refer to Chapter 2, Part 3 of the Revised JICA Guideline for Project Evaluation as for the viewpoints regarding five evaluation criteria

*5: Definition of Recommendation and Lessons Learned
Recommendations: include specific measures, suggestions, and advice on a target project for JICA or those concerned in the implementation agencies
Lessons Learned: can be learned through the experience of a target project and fed back to on-going similar projects or to project finding and planning process in the future