

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 objectivity 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 the following pages.

This year's secondary evaluation first examined the quality of terminal evaluations (primary evaluation) conducted by JICA in fiscal 2004 and 2005. Also, based on the information contained in the terminal evaluation reports, individual projects were evaluated by the working group. In addition, as a new attempt, field studies in relation to seven projects, which were selected based on the results of the past secondary evaluations, were carried out by some members of the Advisory Committee on Evaluation.

In terms of quality of primary evaluation, the result of the secondary evaluation shows that eight out of nine evaluation criteria received more than three points on a scale of five and quality improvement was observed when comparing the secondary evaluation results of fiscal 2004 and 2005 with 2003. Nonetheless, areas that JICA needs to improve were pointed out: for example, the evaluation team composition and the participation of partner countries in evaluation were not sufficiently described in the reports; more figures and tables should be used for a convincing analysis; and measures for improving the quality of evaluation by

overseas offices are required.

With respect to the quality of projects themselves, the average scores for all the five evaluation criteria reached three on a scale of five and the projects in fiscal 2004 and 2005 were graded higher than those in fiscal 2003. In order to improve the quality of future projects, it was recommended that assessment from "suitability as a means" in the evaluation criteria of "relevance" is required and the viewpoint of "cost-effectiveness" in the evaluation criteria of "efficiency" should be enhanced.

As a result of the field studies conducted by some members of the Advisory Committee on Evaluation, the method of secondary evaluation was basically confirmed to be appropriate and effective. At the same time, improvements in the description of contents and methods of terminal evaluation reports and review of the contents of the checklist used for secondary evaluation were recommended in order to further increase effectiveness of secondary evaluation.

Considering the secondary evaluation results examined from the viewpoints of the third party, JICA will take further steps to implement more effective and efficient projects and improve project evaluation. In particular, emphasis will be placed on improving the evaluation capacity of overseas offices. Also, JICA will strive to develop evaluation methods in relation to cost-effectiveness.

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 45 terminal evaluation reports (60 reports if the previous years are included) from various aspects and exercised their ingenuity in conducting secondary evaluation.

Chapter 1 Results of Secondary Evaluation Fiscal 2006

Advisory Committee on Evaluation/
Secondary Evaluation Working Group

1-1 Objectives, Targets, Methods of Evaluation

(1) Objectives

Who should evaluate ODA projects? There might be a number of potential evaluators. If evaluations are performed by stakeholders, it is expected that detailed evaluation in light of circumstances are possible since the evaluators have profound knowledge of the project and region and fully understand the activities and various situations. Also, feedback will more likely fully function, leading to improvements in the project. On the other hand, it could result in lenient evaluations since they may make too much allowance for circumstances, which gives rise to problems in transparency and neutrality. Due partly to the nature of its operation, JICA manages a number of relatively small-scale projects, and for the terminal evaluation alone, the number is around 50 every year. Therefore, JICA, in reality, can not but conduct the internal evaluation; or if not that, it has to seek the assistance of outside stakeholders, such as the supporting committee members, to conduct the evaluation.

Accordingly, as a means of overcoming the expected disadvantages while taking advantage of internal evaluation, objectivity and neutrality can be achieved by conducting internal evaluation thoroughly in compliance with the guidelines and through secondary evaluation by external experts on the results of the internal evaluation. This secondary evaluation does not aim to re-evaluate individual projects but to grasp the general trend of the quality of terminal evaluations and suggest ways for improvement.

The Plan-Do-Check-Act (PDCA) cycle is an effective tool to improve projects constantly. Evaluation corresponds to the Check part of this cycle. If the concept is applied to the process of evaluation, the PDCA cycle of evaluation will be referred to as planning of evaluation, implementation of evaluation, evaluation of evaluations, and improvement of evaluation. In order to diminish the bias of evaluation, it is important to incorporate the views of external examiners; however, in reality, they do not necessarily have to evaluate every single project. At the least, a certain level of transparency and objectivity can be secured if the external examiners' view is incorporated into the Check part of the PDCA cycle.

Evaluation is a series of processes that includes collecting information, conducting analysis, drawing out recommendations/lessons, and compiling reports based on an evaluation framework.

In order to ensure reliability of primary evaluation such as ter-

minal evaluation of projects as in previous years and facilitate the disclosure of easy-to-understand evaluation results, the secondary evaluation in fiscal 2006 was performed with a focus on the following questions.

- a. Evaluation of the quality of primary evaluation
 - Is the primary evaluation qualified enough?
 - Has the quality of primary evaluations improved year by year?
 - What tasks should be carried out to further upgrade the quality?
- b. Evaluation of projects by secondary evaluators based on the reports (i.e. primary evaluation)
 - What is the result of secondary evaluation of the project?
 - Have the evaluation results of projects improved year by year?
 - What are the factors that influence the evaluation results of projects?

(2) Evaluators

There occurs a question about who conducts secondary evaluation. The principle of secondary evaluation refers to whether the evaluation results themselves are convincing, rather than whether they are correct or incorrect. There is no single answer to this question of how evaluation should be carried out, but the answer varies depending on the evaluator's background and the sense of value that affects the evaluation. If numerical targets are set for projects, it is easier to agree as to whether the project purposes have been achieved or not. It is still natural that there are gaps in opinions on the reasons for it and response measures. Even if a secondary evaluator has been provided, there is no guarantee that his/her secondary evaluation result is the utmost and foremost. It is quite probable that results are different when another evaluator conducts secondary evaluation. In such cases, it is safer and more practical to come up with a framework to allow opinions of several secondary evaluators with some level of ability, rather than finding one excellent evaluator.

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

It was fiscal 2003 when JICA launched the secondary evalu-

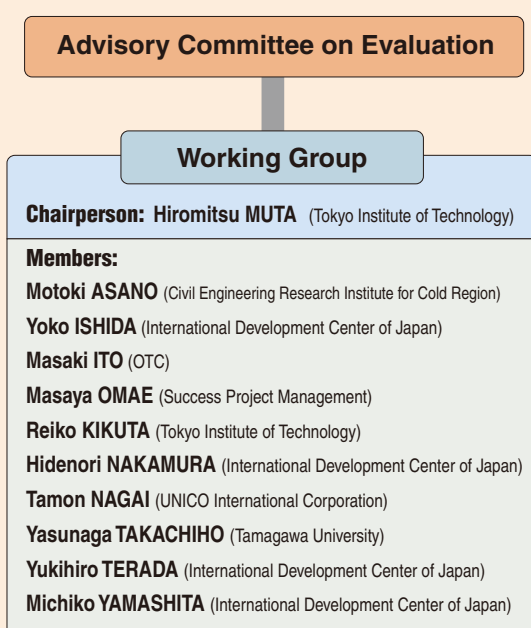
ation on the terminal evaluation by setting up the Working Group. Eight members of the Advisory Committee on Evaluation took charge of the secondary evaluation in fiscal 2003. In fiscal 2004, the Secondary Evaluation Working Group was formed under the Advisory Committee on Evaluation, consisting of six experts and eight JICA staff members (primarily evaluation chiefs of each department). The experts were selected on the basis of recommendations by the Japan Evaluation Society to guarantee the objectivity of the selection. The differences in evaluation tendency between experts and JICA staff were also observed, which leads to the conclusion that there is no significant difference in evaluation tendency between them.

As the development of methodology became complete enough for practical application, in fiscal 2004, the work of the secondary evaluation was subcontracted to the Japan Evaluation Society in fiscal 2005. In fiscal 2006 also, the work was subcontracted to the Japan Evaluation Society. The Society formed an evaluation team comprising 10 members. The members were recruited within the Society for transparency. The secondary evaluation system of fiscal 2006 is illustrated in Figure 4-1.

(3) Evaluation Targets

The secondary evaluation of this year targeted 28 terminal evaluations conducted in fiscal 2004 and 17 terminal evaluations in fiscal 2005. All of these 45 projects were subjects for study. Moreover, for the year-to-year comparison six evaluations in 2004 and nine evaluations in 2003, which had been subject to the previous secondary evaluation, were selected for the secondary evaluation in this year (Appendix 1).

Figure 4-1 Secondary Evaluation System



(4) Evaluation Design and Methods

If it is possible for all the study members to study all the evaluation reports, the mean scores for each evaluation criterion (evaluation viewpoint/criterion) will reflect the opinions of all the evaluators and the results will be 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 different backgrounds and opinions. However, in reality, the workload placed on each evaluator is too enormous to carry out. For example, it takes two to three hours at least for an evaluator to read the report thoroughly and fill in the scores and comments on an evaluation sheet. Including those that had been taken over from the previous year, the number of reports can be more than 60 per year. It would be an excessive burden for the working team, even if it is not impossible.

Thus, in fiscal 2006, each evaluation report was read by four different evaluators. For example, a member of the evaluation working group shared all 60 reports, two shouldered 30 reports each, one member read 24 reports, and six members read 16 reports each. This scheme allows us to treat the judgment criteria of the one key member as the norm of the entire group and adjust the judgment criteria of the other evaluators. Although fairness is more likely with four evaluators rather than one, it is still unavoidable that the tendency of the specific evaluators could affect the results. Theoretically, the scores given by each evaluator can be divided into two parts: true score of the evaluation target (free of personal evaluation bias of the evaluator) and coefficient of evaluation tendency for each evaluator (strictness/leniency coefficient: error tendency of individual evaluator). Accordingly, as in previous years, a method of statistical analysis (analysis of variance) was employed in order to differentiate these two parts so that the evaluation tendency of evaluators was adjusted to obtain the unbiased estimate of evaluation scores that are free of personal evaluation tendency.

The comparative study was conducted year-by-year by sampling the series of the evaluation of a project over years. These projects that had been evaluated repeatedly can serve as so-called “seam allowance” for equating. Using the seam, it is possible to correlate link the secondary evaluation results of fiscal 2005 and 2006. True estimates of the evaluation scores were calculated for fiscal 2005 and fiscal 2006; 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 for two years based on the seam by conversion. Specifically, the results of the secondary evaluation of fiscal 2005 should be converted in such a way so that the mean scores and variances for each fiscal year corresponded to each other. With the proper sampling for the seam, such a simple conversion is sufficient to make a comparison. In this way, evaluation data obtained individually can be processed and analyzed as a large pooled sample through the equating of disconnected evaluation information in various ways using the seam allowance.

(5) Structure of Evaluation Sheet

The secondary evaluation of terminal evaluation has two objectives. One is to evaluate the quality of terminal evaluation and the second is to check the quality of a project using the terminal evaluation.

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

Improvements were made on the evaluation sheet for fiscal 2006 such as by adding evaluation items based on the evaluation results of fiscal 2005. For example, for the criterion of “evaluability,” two viewpoints were set: “evaluability of project purpose” and “evaluability of overall goal.” In response, in place of one viewpoint with regard to the assessment of performance, three viewpoints were added: “measurement of results (outputs),” “measurement of results (project purpose),” “measurement of results (overall goal).” Furthermore, since the criterion of “recommendations/ lessons learned” was rated low in the previous years, the criterion was divided into “recommendations” and “lessons learned” for clearer understanding of the issues and making the evaluation sheet easier to check. The five-point rating scale makes it possible for a year-to-year comparison with the results of fiscal 2005. Table 4-1 shows the changes in evaluation viewpoints from fiscal 2004 to 2006.

The evaluation viewpoints in fiscal 2006 are shown in Table 4-2 and Appendix 2. In the following section, analysis was made based on these evaluation viewpoints. Evaluation was made on the basis of the following five-point scale for rating both view-

points and scoring.

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

1-2 Quality of Terminal Evaluation Examined through Reports

(1) Overview of Evaluation Results

The secondary evaluation in the last fiscal year targeted a total of 45 terminal evaluations (28 evaluations in fiscal 2003 and 17 in fiscal 2004). This fiscal year, 45 terminal evaluations were assessed: 28 in fiscal 2004 and 17 in fiscal 2005, after the last evaluation. The average scores for individual evaluation criteria are shown in Figure 4-2. Among all the criteria, which are “evaluability,” “evaluation framework,” “data collection,” “assessment of performance,” “analysis method,” “five evaluation criteria,” “recommendations,” “lessons learned,” and “reporting,” the scores are relatively high for the criteria of “data collection” for evaluation, “assessment of performance” in analyses, “five evaluation criteria” associated with appropriateness of analysis of DAC’s five evaluation criteria, future “recommendations,” and “lessons learned.” However, the average scores for “evaluability” that asks whether an appropriate evaluation is possible, “analysis method,” and “reporting” are slightly lower. The average score for “evaluation framework,” which concerns the composition of the evaluation team and the degree of the partner country’s participation in evaluation, failed to reach 3.0, whereas the average

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

Fiscal Year	2004		2005		2006	
	Viewpoints	Rating	Viewpoints	Rating	Viewpoints	Rating
Evaluability	4 (3-point scale)	10-point scale	4 (5-point scale)	5-point scale	6 (5-point scale)	5-point scale
Evaluation Framework	4 (3-point scale)	10-point scale	3 (5-point scale)	5-point scale	2 (5-point scale)	5-point scale
Data Collection	5 (3-point scale)	10-point scale	4 (5-point scale)	5-point scale	4 (5-point scale)	5-point scale
Assessment of Performance	4 (3-point scale)	10-point scale	4 (5-point scale)	5-point scale	6 (5-point scale)	5-point scale
Analysis Method	3 (3-point scale)	10-point scale	3 (5-point scale)	5-point scale	3 (5-point scale)	5-point scale
Evaluation (Result by DAC’s Five Criteria)	7 (3-point scale)	10-point scale	6 (5-point scale)	5-point scale	6 (5-point scale)	5-point scale
Recommendations	4 (3-point scale)	10-point scale	3 (5-point scale)	5-point scale	3 (5-point scale)	5-point scale
Lessons Learned	4 (3-point scale)		3 (5-point scale)		3 (5-point scale)	5-point scale
Reporting	4 (3-point scale)	10-point scale	3 (5-point scale)	5-point scale	3 (5-point scale)	5-point scale
General Criteria for Good Evaluation	4 (3-point scale)	10-point scale				
Evaluation of the Project: Relevance		10-point scale	3 (5-point scale)	5-point scale	3 (5-point scale)	5-point scale
Evaluation of the Project: Effectiveness		10-point scale	2 (5-point scale)	5-point scale	2 (5-point scale)	5-point scale
Evaluation of the Project: Efficiency		10-point scale	2 (5-point scale)	5-point scale	3 (5-point scale)	5-point scale
Evaluation of the Project: Impact		10-point scale	3 (5-point scale)	5-point scale	3 (5-point scale)	5-point scale
Evaluation of the Project: Sustainability		10-point scale	5 (5-point scale)	5-point scale	5 (5-point scale)	5-point scale
Evaluation of the Project: Overall Evaluation		10-point scale				

Table 4-2 Secondary Evaluation Criteria

<p>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</p>
<p>II. Key Evaluation Criteria</p> <p>1. Criterion: Evaluation Framework Viewpoints: Time Frame of Evaluation Study Evaluation Team Composition—Impartiality and Specialty Level of Counterpart Participation</p> <p>2. Criterion: Data Collection Viewpoints: Evaluation Questions Appropriateness of Data Collection Methods and Data Sources Data/Information Sources Sufficiency of Data/Information Obtained</p> <p>3. Analysis</p> <p>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 Relationships—Logic of Project Design •Examination of Quantitative Causal Relationships—Before and After</p> <p>3.2 Criterion: Analysis Method Viewpoints: •Objective Analysis •Holistic Analysis •Analysis of Promoting and Impeding Factors</p> <p>4. Criterion: DAC’s Five Evaluation Criteria Viewpoints: •Relevance •Effectiveness •Efficiency •Impact •Sustainability •Conclusion</p> <p>5. Recommendations/Lessons Learned</p> <p>5.1 Criterion: Recommendations Viewpoints: •Sufficiency of Recommendations •Relevance and Credibility of Recommendations •Usability of Recommendations</p> <p>5.2 Criterion: Lesson Learned Viewpoints: •Sufficiency of Lessons Learned •Relevance and Credibility of Lessons Learned •Usability of Lessons Learned</p> <p>6. Criterion: Reporting Viewpoints: •Presentation/Legibility and Clarity •Utilization of Tables and Figures •Presentation of Primary Data</p>
<p>III. Project Evaluation Based on the Written Report (DAC’s Five Criteria)</p> <p>1. Criterion: Relevance Viewpoints: •Validity •Necessity •Appropriate Approach</p> <p>2. Criterion: Effectiveness Viewpoints: •Achievement Level of Project Purpose •Causal Relationships between Outputs and Project Purpose</p> <p>3. Criterion: Efficiency Viewpoints: •Clear Input Cost •Cost-benefit Performance •Appropriate Implementation Process</p> <p>4. Criterion: Impact Viewpoints: •Achievement Level of Impact •Logics on Causal Relationships of Impact •Unanticipated Impact (Both Positive and Negative)</p> <p>5. Criterion: Sustainability (Post-JICA’s Cooperation) Viewpoints: •Mechanism of Securing Sustainability •Level of Sustainability •Organizational Sustainability •Technological Sustainability •Financial Sustainability</p>

scores for eight other evaluation criteria out of nine, except for “evaluation framework,” are 3.0 or higher.

As for the distribution of scores, as shown in Figure 4-3, many are distributed between 2.5 and 4.49. Many projects scored 3.0 or higher for “evaluability,” “data collection,” “assessment of performance,” “five evaluation criteria,” “recommendations,” and “lessons learned.” For “evaluation framework” and “analysis method,” half of the projects scored less than 3.0 and the other half 3.0 or higher.

In sum, it can be concluded that the quality of terminal evaluations belongs to the higher level than “medium” on the grading scale.

(2) Evaluation Results by Criterion

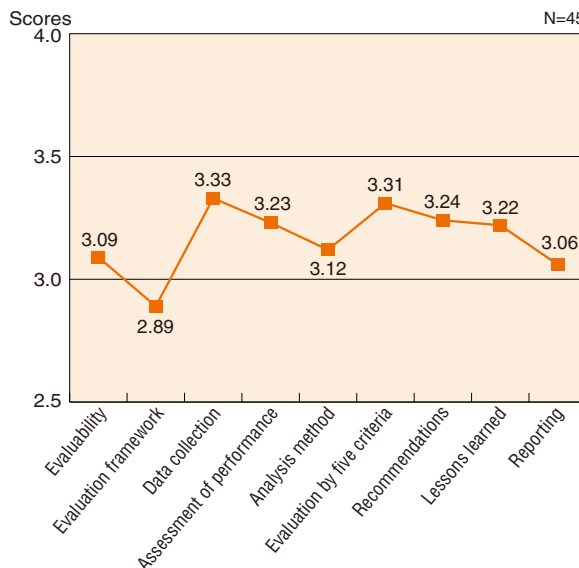
The viewpoints of each evaluation criterion were rated, and qualitative data was collected in the forms of comments of the evaluators that were written in the additional boxes on the sheet. We will summarize the current conditions and issues of the quality of terminal evaluation by criterion based on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

1) Evaluability

“Evaluability” is a criterion that asks about 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 the beneficiary of the project, was set clearly and properly
- Evaluability of project purpose: whether the indicators and

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



specific target values are clearly defined for each output and project purpose so that they can be used to measure the level of the 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 the project achievement
- Logic of project design: whether the PDM used for the evaluation describes a clear and realistic logic flow from Overall goal – Project Purpose – Outputs – Inputs, considering important external assumptions
- Project monitoring: whether monitoring of outputs, activities, and inputs was regularly conducted, and the information including statistical data was accumulated during project implementation

The score results (average) reveal gaps in quality of the evaluations for each viewpoint. For example, the scores for “target group,” “logic of project design,” and “project monitoring” are 3.1 or higher, securing the “average” level or higher in the grading scale. Among them, “target group” scored 3.3 on average, which is higher than other viewpoints. On the other hand, “evaluability of overall goal” scored 2.8 on average and failed to

achieve the “average” level. The specific item, “evaluability of overall goal,” is a new viewpoint that was added in 2006. The low total scores are likely to be attributed to the fact that there were not a few projects with a weak causal relationship between the overall goal and project purpose.

The projects with high scores have tendencies: they logically and clearly set the project purposes and indicators and incorporated baseline studies, or were appropriately designed with thorough preparations based on past experiences. Also, high scores were given to those projects that conducted monitoring using PDM that had been improved at the time of mid-term evaluation, and to one that carefully planned and implemented monitoring and collected data necessary for evaluations every year. On the other hand, projects whose indicators to measure the achievement of the goals were not set, or were too abstract/inappropriate to conduct evaluation, and projects in which PDM was less accepted and not utilized for monitoring turned out to have low ratings.

2) Evaluation Framework

“Evaluation framework” refers to the evaluators of terminal evaluation. This criterion consists of two viewpoints.

- Evaluation team composition: whether the evaluation team

Figure 4-3 Distribution of Scores by Evaluation Criterion

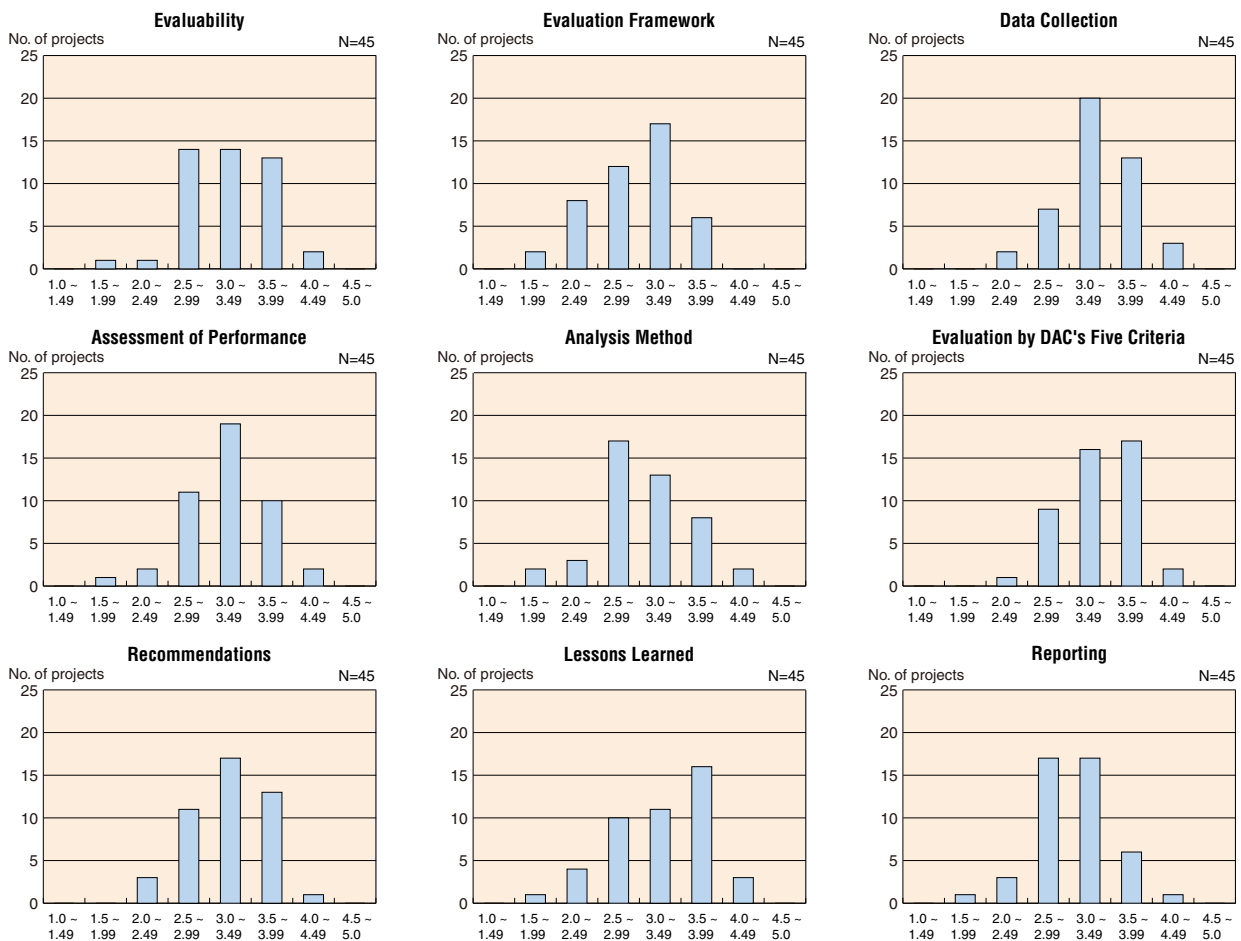
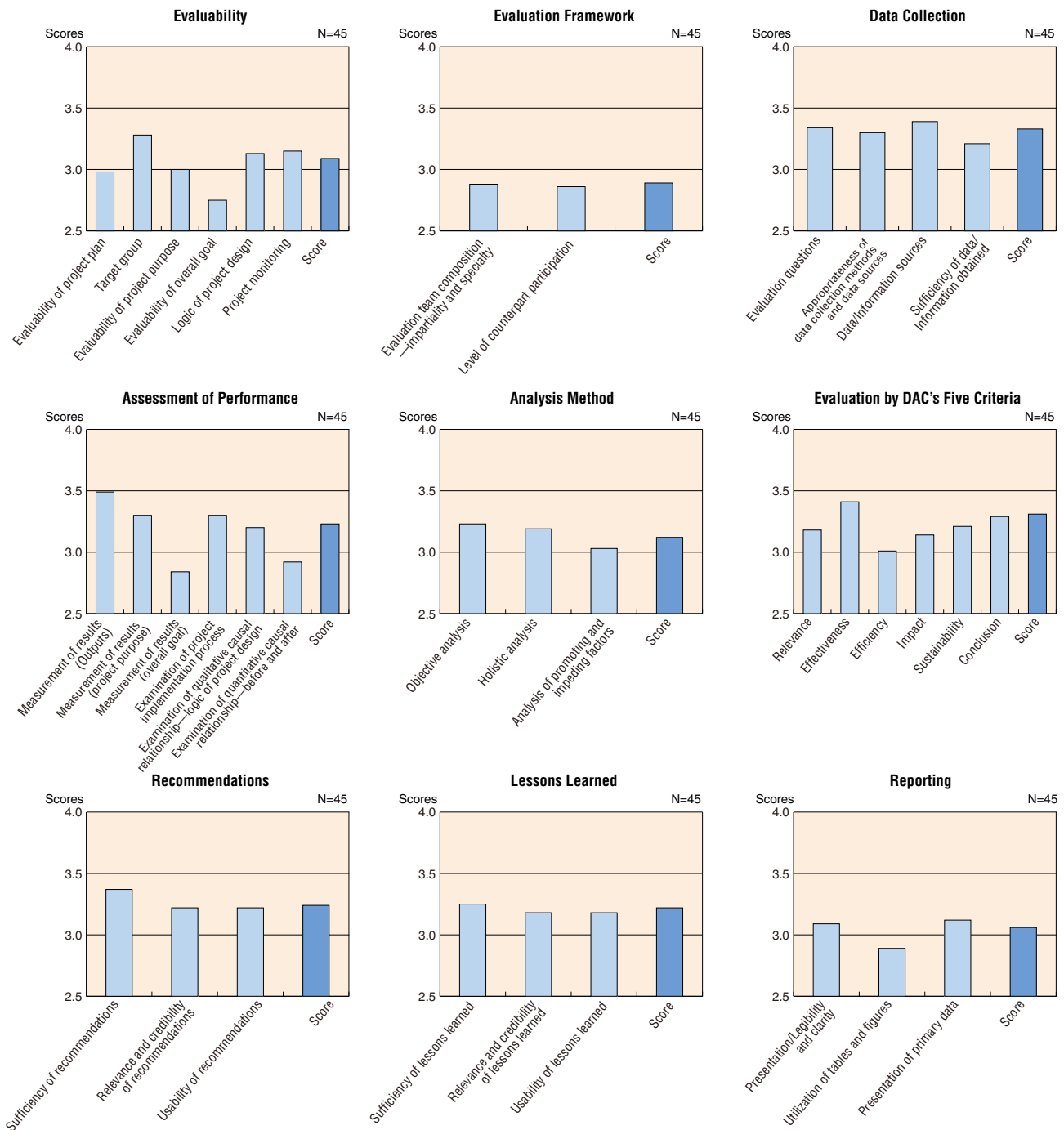


Figure 4-4 Score Results for Viewpoints under Each Evaluation Criterion (Average)



members of the project are qualified enough to conduct professional and impartial evaluations

- Level of counterpart participation: whether the counterparts in the developing country participated sufficiently in the evaluation as evaluators

The average scores for these viewpoints all resulted in 2.9 points, and so did not clear the “medium” level, and are lower than the viewpoints of other evaluation criteria.

Regarding the evaluation team composition, evaluation teams consisted of JICA staff members, evaluation analysts, and other

members. There is a case where project stakeholders such as members of the domestic support committee, consultants, or dispatched experts join the team as members. In another case, the team consisted of two members: one JICA staff member and one local consultant. Although such team formation may gain considerable specialization, it is somewhat difficult to secure neutrality and impartiality. Such cases may lead to the low scores for the composition of evaluation team.

“Level of counterpart participation” checks to what extent the partner country actually participated in the project evaluation process. In the case of highly rated projects, they prepared

joint evaluation and not only identified the names and specialty areas of evaluators from the partner country, but also guaranteed the objectivity of those evaluators and provided detailed information on the method of joint evaluation as well as the number of meetings. On the other hand, in the cases with low scores, joint evaluation teams were not formulated; the report did not reveal the names of team members from the partner country and did not describe how the evaluation was performed even though it mentioned the evaluation, which made it hard to determine the level of participation of the partner countries in the evaluations.

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 set in line with the evaluation purposes appropriately so that they can 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 the data/information are explained adequately in the evaluation report
- Data collection: whether the data/information was sufficient to answer the evaluation questions in terms of both quality and quantity

The viewpoints on “data collection” show little variance in the quality of evaluation, with all the average scores being over 3.2 and attaining the “medium” level. Thus, the scores are relatively higher than viewpoints on the other evaluation criteria. The average score of 3.4 on “data/information sources” is relatively higher than those on the other evaluation viewpoints.

It is reported that many projects studied collected data appropriately. There was a project in which stakeholders had such a strong awareness about monitoring that they accumulated data through daily information updates in preparation for evaluations. Also, there was a project in which evaluation questions were appropriately narrowed down, data were collected from wide sources covering existing documents, relevant authorities, project stakeholders, training participants and host companies, etc, and the information sources were clearly identified. On the contrary, some projects reportedly did not collect sufficient data. For example, the data sources were biased: in one project the data sources were limited to stakeholders, while in another data was not collected to assess the achievement of the overall goal.

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 achieve-

ment 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 to achievement of outputs, project purpose, and overall goal are identified
- Examination of qualitative causal relationships—logic of project design: whether the logic of the project design was thoroughly verified, through which impeding and/or promoting factors to achievement of 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 the beneficiaries have resulted from the project interventions

In the 2006 evaluation, a slight revision was made to “assessment of performance.” That is, measurement of results was divided into three categories: outputs, project purpose, and overall goal. Also, viewpoints on causal relationships were better specified by paraphrasing them as “qualitative causal relationship” and “quantitative causal relationship.”

The evaluation scores for the viewpoints in “assessment of performance” show large variance in the quality of evaluations. Among the viewpoints, the average scores for “measurement of results (outputs),” “measurement of results (project purpose),” “examination of project implementation process,” and “examination of qualitative causal relationships” are all 3.2 or higher. The viewpoint, “measurement of results (outputs)” scored the highest of all viewpoints at 3.5. On the other hand, the average scores for “measurement of results (overall goal)” and “examination of quantitative causal relationships” are less than 3.0, thus failing to reach the “medium” level.

In the projects with the high scores for evaluation, causal relationships with the goal achievement and the current conditions about important external factors in line with PDM, or the details of successful and unsuccessful factors in relation to the achievement level of outputs were separately understood. On the other hand, some projects did not fully assess results due to discrepancies between the overall goal and the project purpose, and other projects assessed the current conditions qualitatively, but did not get to do so quantitatively, resulting in a lack of quantitative data. It is assumed these caused low scores for “measurement of results (overall goal)” and “examination of quantitative causal relationships.”

5) Analysis Method

In “analysis method,” evaluation is evaluated based on the following three viewpoints to check how analysis is performed.

- Objective analysis: whether the data was objectively analyzed based on a series of scientific discussions, and an effort was made to quantify the data where feasible
- Holistic analysis: whether the data interpretation was drawn by 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 the project implementation process (such as project management)

The average score results for all the viewpoints achieved 3.0 or higher (average) at the “medium” level. The projects receiving high scores include one that carried out monitoring at the implementing stage and implemented holistic analyses based on the data accumulated in the monitoring over a number of years; and another that objectively analyzed outputs, process, and logic from holistic standpoints, though the analysis was qualitative, before drawing a conclusion. Another highly-rated project performed analyses appropriately using quantitative data, questionnaires, and interviews. On the other hand, low scores were rated due to insufficient analyses in terms of objectivity and holistic approach; specifically, objectivity suffered when analyses were based on the limited information from project stakeholders and the holistic approach was compromised when data were obtained from questionnaires alone. Furthermore, some projects performed insufficient analyses on promoting and impeding factors although those factors were present, and other projects conducted analyses that lacked logic.

6) Evaluation by Five Criteria

“Evaluation” means to evaluate based on six evaluation viewpoints: DAC’s five evaluation criteria plus “conclusion,” which is to check whether the conclusion was drawn based on holistic viewpoints on the basis of the concerned five evaluation criteria.

- Relevance: whether 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.) were sufficiently covered
- Effectiveness: whether perspectives for evaluation of “Effectiveness” (achievement level of project purpose, causal relationships between outputs and project purpose, etc.) were sufficiently covered
- Efficiency: whether perspectives for evaluation of “Efficiency” (comparison with other similar projects through cost analysis, cost-effectiveness analysis, etc.) were sufficiently covered
- Impact: whether perspectives for evaluation of “Impact” (achievement level of overall goal, causal relationships between project purpose and overall goal) were sufficiently covered
- Sustainability: whether the perspective for evaluation of “Sustainability” (mechanism of securing sustainability and outcomes to be produced in terms of policies and systems, orga-

nizational and financial aspects, technical aspects, socio-culture, and environment) were sufficiently covered

- Conclusion: whether the conclusion was drawn based on holistic viewpoints on the basis of the five evaluation criteria.

The average score results for all the viewpoints received 3.0 or higher, exceeding the “medium” level, yet showing a variance in evaluations of the viewpoints. Among the viewpoints, the average score for “effectiveness” is the highest at 3.4 and is relatively higher than viewpoints on other evaluation criteria.

For “efficiency,” the average score was the lowest of the six viewpoints at 3.0. “Efficiency” was rated the lowest in the secondary evaluation for fiscal 2005 as well. This criterion questions whether perspectives (comparison with other similar projects through cost analysis, cost effectiveness, etc.) are sufficiently covered. Many projects did not mention the comparison with similar projects and cost-effectiveness, and did not provide adequate information concerning operating costs, all of which led to the low score.

7) Recommendations

The criterion, “recommendations,” concerns the following three viewpoints.

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

The average scores for these viewpoints are 3.2 or higher, securing relatively high evaluation considering a small variance of the quality in evaluations. Among them, “sufficiency of recommendations” is high at 3.4.

Many projects appropriately assessed the achievement of results and provided holistic recommendations with positive and negative factors logically summarized; provided practical recommendations through analyses of the project outputs; and provided recommendations based on evaluation analysis. These projects received high scores. On the other hand, some projects drew out recommendations from other than evaluation processes, and did not provide clear bases for recommendations without presenting clear relationships between evaluation results and recommendations. In some projects, impeding factors were not reflected on recommendations because analysis of impeding factors was insufficient, and the background of recommendations was not clear because recommendations were presented in an itemized list. These projects received low scores.

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 the information obtained through the process of data analysis and interpretation and as a result, the lessons learned are objective and convincing
- Usability of lessons learned: whether the lessons are generalized and conceptualized so that they are widely applicable

The average scores are 3.2 or higher, which is relatively high, with a small variance in the quality of evaluations.

Many projects presented reasonable lessons by identifying promoting and impeding factors from fully analyzed results and evaluation processes, and provided useful lessons for other similar projects with specific descriptions. In some cases, however, no specific directions were provided for what to do and how to do it, so that few could be applied to other projects because of insufficient analysis of impeding factors, or some lessons were confined to superficial remarks.

9) Reporting

“Reporting” covers the three viewpoints shown below.

- Presentation/legibility and clarity: whether the evaluation report is simple and clear, and understandable to readers—in light of the 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 those on targets and results of interviews and questionnaires or sources are presented properly in the report

The scores for “presentation/legibility and clarity” and “presentation of primary data” are 3.1 or higher, while “utilization of tables and figures” is 2.9, failing to reach the average level.

The reports of some projects were well written and easy to read; for example, the correspondence between outputs/purpose and indicators was clearly demonstrated, evaluations were performed in line with PDM, tables and figures were effectively used, and the description was plain and logical. On the other hand, however, some reports remained too general, without provision of PDM, and too many pages were devoted to the list of participants for seminars and workshops without clear description as to how the evaluation analysis was done. In fiscal 2006, several terminal evaluation reports were written in English and compiled by overseas offices, but they generally lacked evaluation analyses, provided too many descriptions about materials other than evaluations, which received low scores in quality in the end.

(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 were evaluated from the nine evaluation criteria: “evaluability,” “evaluation framework,” “data collection,” “assessment of performance,” “analysis method” “evaluation by five criteria,” “recommendations,” “lessons learned,” and “reporting.” The overall quality of the terminal evaluations was ranked 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 and the lowest is one, and the medium level is three as the average. We selected the top four cases of evaluations and the worst four cases, while giving consideration to the distribution of the overall scores. The scores for nine criteria of these eight evaluations are shown respectively on the web graphs in Figure 4-5 and Figure 4-6. Table 4-3 shows the average scores and the differences in the average scores for evaluation criteria of the top four evaluations and the last four evaluations.

As clearly observed in Figure 4-5 and Table 4-3, the average scores of the top four evaluations are quite high at 3.8 or higher for “data collection,” “assessment of performance,” “analysis method,” “criteria by five evaluation,” and “lessons learned.” In particular, the score for “evaluation by five criteria” is high at 4.0. In specific terms, the factors contributing to such high scores are: data sources are clear, appropriate data were sufficiently collected, the implementation process, performance and effects of projects were fully assessed and examined from the qualitative and quantitative aspects, collected data were objectively analyzed from various aspects and promoting and impeding factors to the onset of effects were analyzed, evaluations were conducted covering necessary prospective with respect to five evaluation criteria, and useful lessons that are objective and convincing and can be used for similar projects were drawn from the information on impeding and promoting factors obtained during the evaluation processes.

On the other hand, as evidenced in Figure 4-6 and Table 4-3, for the worst four projects, there is a tendency where the scores for “evaluability,” “evaluation framework,” “assessment of performance,” and “reporting” are relatively low. As indicators for the project purpose and overall goal are vague and lack logic, it was difficult to assess performance, which resulted in an obscure basis for determining performance and weak quantitative analysis. There were some reports that neither contained primary data nor

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

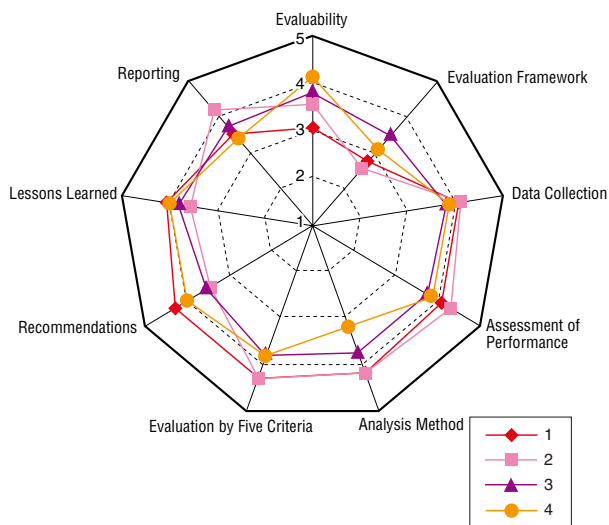


Figure 4-6 The Worst 4 Projects for its Quality of Terminal Evaluation

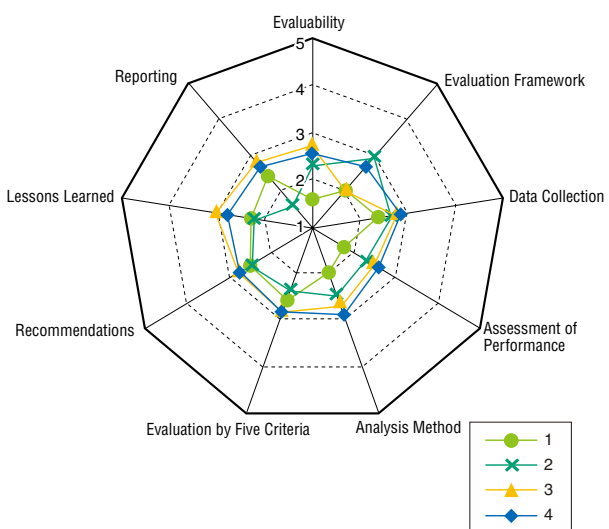


Table 4-3 Score Results of the Top 4 and Worst 4 Projects (Average)

Evaluation Criteria	Average Scores		Difference in Average Scores
	Top 4 projects	Worst 4 projects	
Evaluability	3.58	2.29	1.29**
Evaluation Framework	3.01	2.38	0.63
Data Collection	3.90	2.62	1.29**
Assessment of Performance	3.97	2.26	1.71**
Analysis Method	3.81	2.51	1.30**
Evaluation by Five Criteria	4.03	2.63	1.40**
Recommendations	3.77	2.58	1.19**
Lessons Learned	3.82	2.58	1.24**
Reporting	3.69	2.33	1.36**

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

provided necessary information. These factors caused the low scores for the quality of the terminal evaluations.

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

The four evaluation reports rated as being high in quality are listed in Table 4-4, and those rated as being poor in quality in Table 4-5.

The project that obtained the highest score for its terminal evaluation is Project on Establishment of Control Capacity for Industrial Wastewater and Waste in Argentina. In the basin area of the Matanza River, which runs through the Buenos Aires district, soil and groundwater contamination were caused by domestic and industrial wastewater, calling for urgent action to control toxic materials. However, the National Water Institute of Argentina did not possess technology, knowledge, and experience to fully address the issue. Thus, the project was implemented with the purpose of empowering the National Water Institute to develop a foundation for chemical analysis, contamination assessment, and cleaner production in order to tackle the industrial pollution issues. Meanwhile, since there was a change in the environment surrounding the project, such as the national drastic economic crisis, the project plan was revised so as to enable the project to carry out its activities without relying on the Matanza Project, which had been premised on IDB support. The technology was smoothly transferred and the project purpose will most likely be achieved. Furthermore, a positive impact was observed; the project set up a new department called the Cleaner Production and Sustainable Consumption Unit within the Environment and Sustainable Development Agency, which increased the understanding of cleaner production in the industrial sector, and thus, production process improvements have started.

With regard to quality of the terminal evaluation, all the evaluation criteria of “data collection,” “assessment of performance,” “analysis method,” “evaluation by five criteria,” “recommendations,” and “lessons learned” scored more than four points. The terminal evaluation of this project was determined to be quite qualified for the following reasons.

In the aforementioned evaluation, sufficient data were collected. The assessment of achievement level based on the indicators and the examination of the implementation process as to the response to changes in important external factors are clearly described. Also, objective analyses from various aspects are performed, including impeding factors and the background of the revision of PDM. Evaluations were fair and valid based on the achievement level. Furthermore, the conclusion is logical enough, and the recommendations are described in a concrete and appropriate way and suitable for the project. Lessons learned are also

Table 4-4 Terminal Evaluations of Good Quality

Country	Project Name	Total Score	Fiscal Year of Evaluation
1 Argentina	Project on Establishment of Control Capacity for Industrial Wastewater and Waste	3.80	2004
2 Thailand	Project for Development of Trauma Center Complex	3.78	2004
3 Philippines	Water Buffaloes and Beef Cattle Improvement Project	3.68	2005
4 Ghana	Improvement of Educational Achievement in Science, Technology and Mathematics (STM) in Basic Education	3.67	2004

Table 4-5 Terminal Evaluations of Poor Quality

Country	Project Name	Total Score	Fiscal Year of Evaluation
1 Thailand	Project on Local Management Cooperation in Thailand	2.15	2004
2 Viet Nam	Strengthening of National Institute of Veterinary Research	2.36	2004
3 Myanmar	Project for Primary Health of Mothers and Children	2.66	2005
4 Cambodia	The Capacity Building for the Forestry Sector	2.69	2004

described in a specific and detailed manner.

The worst project for its terminal evaluation quality is Project on Local Management Cooperation in Thailand. As decentralization progressed in Thailand, the capacity development of local governments and coordination among them were increasingly sought after to meet the needs of the community. This project thus aimed to formulate guidelines for cooperation among local governments and promote such cooperation. The guidelines are almost completed and discussions for the revision of laws are underway to promote cooperation among local governments.

Looking at the quality of the terminal evaluation, however, the scores for every evaluation criterion are less than 2.6 and the score for “assessment of performance” is less than two points. The quality of terminal evaluation of this project was rated low for the following reasons.

The evaluation standards are unclear partly because PDM and the evaluation grid had not been developed. Data was collected solely from the questionnaires, resulting in insufficiency, and the contents of questions are irrelevant. Assessment of performance was based on the questionnaires alone, no referrals were made to the results of the survey; consequently the performance was assessed without a clear logic. Though a quantitative analysis was attempted, the analysis based on questionnaire surveys alone lacks objectivity. Because project activities were not fully described, it is somewhat unclear as to what and how it was evaluated. By not following the JICA Guidelines in general, the evaluation turned out to be weak in objectives and reasoning.

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

The Secondary evaluation of terminal evaluations has been carried out since fiscal 2003, targeting 38 terminal evaluations conducted in fiscal 2002, 38 in fiscal 2003, 45 in fiscal 2004, and 17 in fiscal 2005, besides those conducted this fiscal year (FY2006). The project evaluations should guarantee high quality of study in order to convey accurately the outcomes of implemented projects to the readers. We will take a look at how the quality of terminal evaluation has changed over the years. The ter-

minal evaluations subject to secondary evaluation in fiscal 2002 were not included in this year-to-year analysis of quality changes because many of the evaluation criteria and viewpoints were far different from those subject to secondary evaluation in fiscal 2003 and afterwards.

The evaluators of secondary evaluation are different every year. Even though evaluation criteria are the same, evaluation viewpoints are slightly different over the years. Rating scales also changed; a 10-level rating scale was adopted in fiscal 2004, whereas a five-level rating scale was used in fiscal 2005 and fiscal 2006. Thus, considering the differences in evaluation standards, the evaluation results of projects that were evaluated twice were used to convert the evaluation scores to unify the evaluation scale. In other words, the evaluation scores of fiscal 2005 were converted to the scale of fiscal 2006. The scores of the terminal evaluations conducted in fiscal 2004 were first converted to the scale of fiscal 2005 and the obtained scores were then converted to the scale of fiscal 2006. It must be noted that the terminal evaluations subject to secondary evaluation twice adopted the original scores, not the converted scores.

Furthermore in fiscal 2004 and fiscal 2005, the recommendations and lessons learned were lumped together as a single criterion, while, in fiscal 2006, they were separated into respective criteria, “recommendations” and “lessons learned.” Thus, for the results of secondary evaluation conducted in fiscal 2004 and fiscal 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 the respective percentage 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, and 17 projects in fiscal 2005 were obtained by evaluation criterion, which are shown in Figure 4-7 and Table 4-6. As for the changes in the average scores between 2003 and 2004, the scores of 2004 are significantly higher statistically for “data collection,” “evaluation by five criteria,” “recommendations,” “lessons learned,” and “reporting”. Regarding the differences between fiscal 2003 and fiscal 2005, the average scores

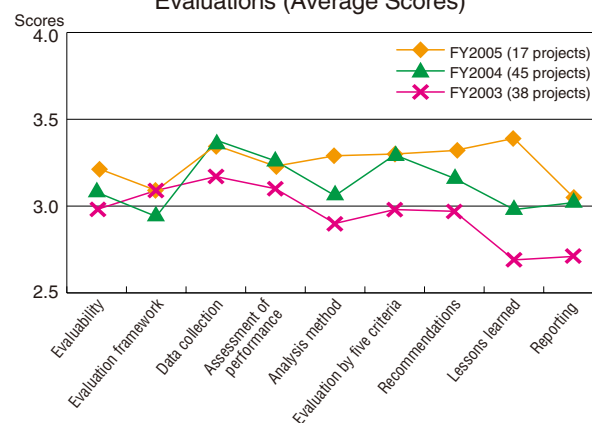
Figure 4-7 Year-to-year Changes in the Quality of Evaluations (Average Scores)

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

Evaluation Criteria	Average Scores			Difference in the Average between Years		
	FY2003(A)	FY2004(B)	FY2005(C)	(B)-(A)	(C)-(A)	(C)-(B)
I Preconditions for Conducting Appropriate Evaluation						
Evaluability	2.98	3.08	3.22	0.10	0.24	0.14
II Key Evaluation Criteria						
Evaluation Framework	3.09	2.94	3.09	-0.15	0.00	0.15
Data Collection	3.17	3.38	3.35	0.21*	0.18	-0.03
Assessment of Performance	3.10	3.26	3.23	0.16	0.13	-0.03
Analysis Method	2.90	3.06	3.30	0.15	0.39**	0.24
Evaluation by Five Criteria	2.98	3.30	3.30	0.32**	0.33*	0.00
Recommendations	2.97	3.16	3.32	0.19*	0.36**	0.16
Lessons Learned	2.70	2.98	3.40	0.29*	0.70**	0.41*
Reporting	2.71	3.02	3.05	0.31**	0.34**	0.03

* 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.

for “analysis method,” “evaluation by five criteria,” “recommendations,” “lessons learned,” and “reporting” of fiscal 2005 are significantly higher statistically. Due to the limited number of samples (17 projects for fiscal 2005) it cannot be generalized, still it is fair to say that the average score for “lessons learned” of fiscal 2005 is statistically higher than those of fiscal 2004 and the scores of fiscal 2005 tend to be high in general. Based on these facts, the quality of terminal evaluation seems to have improved gradually from fiscal 2003 to fiscal 2005, although the difference between fiscal 2004 and fiscal 2005 is not statistically significant.

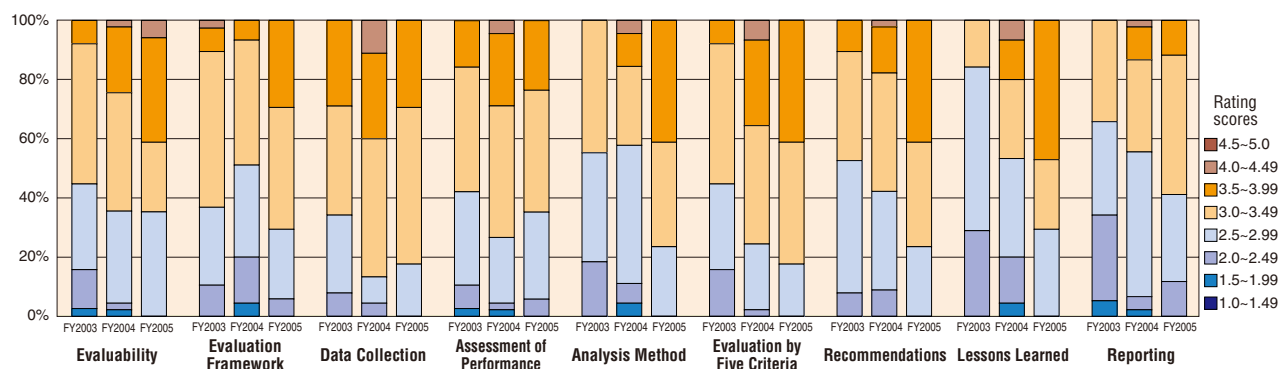
Year-to-year changes in the quality of terminal evaluation were examined last year, too (See p125, Annual Evaluation Report 2005). There is a difference in the average scores of fiscal 2003 and 2004 between last year and this year. This is attributed to the facts that the scores for each criterion were converted to this year’s scale and that the number of projects studied in fiscal 2004 was only 17, while the number increased this year to 45, adding 28.

Last year’s secondary evaluation concluded that the quality of terminal evaluations of fiscal 2004 was higher than that of fiscal 2003. The same conclusion was reached this year. The number of projects of fiscal 2004 for which last year’s (fiscal 2005) secondary evaluation was conducted was 17, but it increased to 45 in fiscal 2006, thus attaining more reliable results. Based on the

above results, it is confirmed that the quality of terminal evaluations of fiscal 2004 improved compared to the previous year.

Next, look at the changes in distribution of the scores for each evaluation criterion. For each, the scores ranging from 1.0 to 5.0 were divided into eight groups of 0.5-point increments. The bar chart, Figure 4-8, illustrates the percentage of reports with scores categorized into the eight groups. If the bar for fiscal 2003 and the one for 2004 are compared, it is clear that in 2004, the percentage of the scores in the 1.0 and 2.0 ranges decreased from those in fiscal 2003 in many criteria, while the percentage of the scores in the 3.0 and 4.0 ranges increased. As for the criteria of “evaluability,” “data collection,” and “assessment of performance,” although there are no statistical differences for average scores by years, the percentage of the reports with scores between 1.0 and 2.9 decreased, while the percentage of scores over 3.5 increased. For “analysis method,” in fiscal 2004 the percentage of the scores between 2.5 and 2.99 as well as those over 3.5 increased and some achieved more than 4.0 points. In fiscal 2005, if compared to 2003 and 2004, the percentage of the scores lower than 2.5 tends to decrease, while the scores over 3.5 increase. From the above results, it can be said that the quality of terminal evaluations of fiscal 2004 and fiscal 2005 improved compared to the one in 2003.

Figure 4-8 Changes in the Quality of Terminal Evaluations (Distribution of Scores, 2003-2005)



(5) Evaluation by JICA Headquarters and Evaluation by 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, which has 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 projects of 2004 were ones carried out by overseas offices and three were selected in 2005.

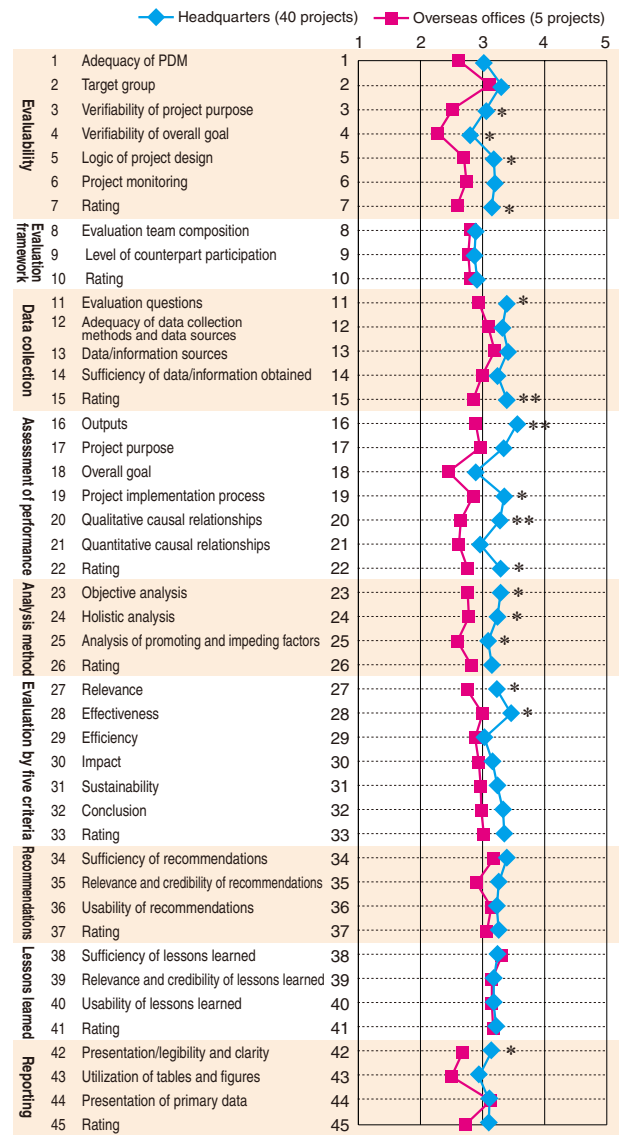
Figure 4-9 compares the scores of the terminal evaluations conducted by JICA headquarters to overseas offices.

Although the number of projects for overseas offices is limited to five, it is clear that the scores of overseas offices tend to be lower in every evaluation criterion than the ones conducted by the headquarters.

In the criterion of “evaluability,” the scores for both “verifiability of project purpose,” “verifiability of overall goal,” and “logic of project design” indicate statistically significant differences. The scores for overseas offices are rated lower than those for headquarters, most of which remain in the 2.0 range. In the “data collection,” statistically significant differences are found for “evaluation questions” and “ratings,” and the scores rated by overseas offices are lower than the ones rated by the headquarters. A statistically significant difference is found for “assessment of performance,” the scores for “outputs,” “implementation process,” and “qualitative causal relationship,” and the scores are again rated lower for overseas offices than the headquarters. As for “analysis method,” the scores for “objective analysis,” “holistic analysis,” and “analysis of promoting and impeding factors” are statistically and significantly different, and the scores rated by overseas offices are lower. None of the scores of overseas offices for “assessment of performance” and “analysis” reached 3.0 points. With respect to “five evaluation criteria,” the scores for “relevance” and “effectiveness” are statistically and significantly different, and the scores rated by overseas offices are lower than those rated by the headquarters. The scores for “presentation/eligibility and clarity” of the criterion of “reporting” show a statistically significant difference, and the scores rated by overseas offices are lower than those by the headquarters.

When looking at the ranking of total scores of evaluation reports compiled by overseas offices, one report is ranked in the sixth place, but the remaining four reports are among the bottom 11 projects. It is therefore clear that the quality of the terminal evaluation reports compiled by overseas offices is relatively low. The reason for the low quality of the terminal evaluation reports of the projects managed by overseas offices may be that evaluations were not performed in line with PDM, which led to a weak logical understanding about the project evaluation and insuffi-

Figure 4-9 Quality of Evaluations Conducted by the Headquarters and Overseas Offices



*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.

cient assessment of performance and analyses.

However, with respect to “recommendations” and “lessons learned,” the scores of overseas offices for most of the evaluation criteria are in the 3.0-3.9 range, attaining the same level of evaluation for the headquarters. “Recommendations” and “lessons learned” are drawn out from the implementation process and achievements of projects; however, it is important to draw out recommendations and lessons based on the situations of not only the projects, but also partner countries. Since overseas offices are more familiar with conditions in the partner countries, this may have contributed to an increase in the “average” level of assessment on “recommendations” and “lessons learned.” High quality evaluations will be possible if overseas offices try to conduct evaluation in line with JICA Guidelines (2004) and use the advantages of overseas offices.

(6) Implementation of Ex-ante Evaluation

JICA has worked to establish a consistent evaluation system from ex-ante to ex-post along 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 and the appropriateness of a project plan based on the expected effects prior to the launch of the project.

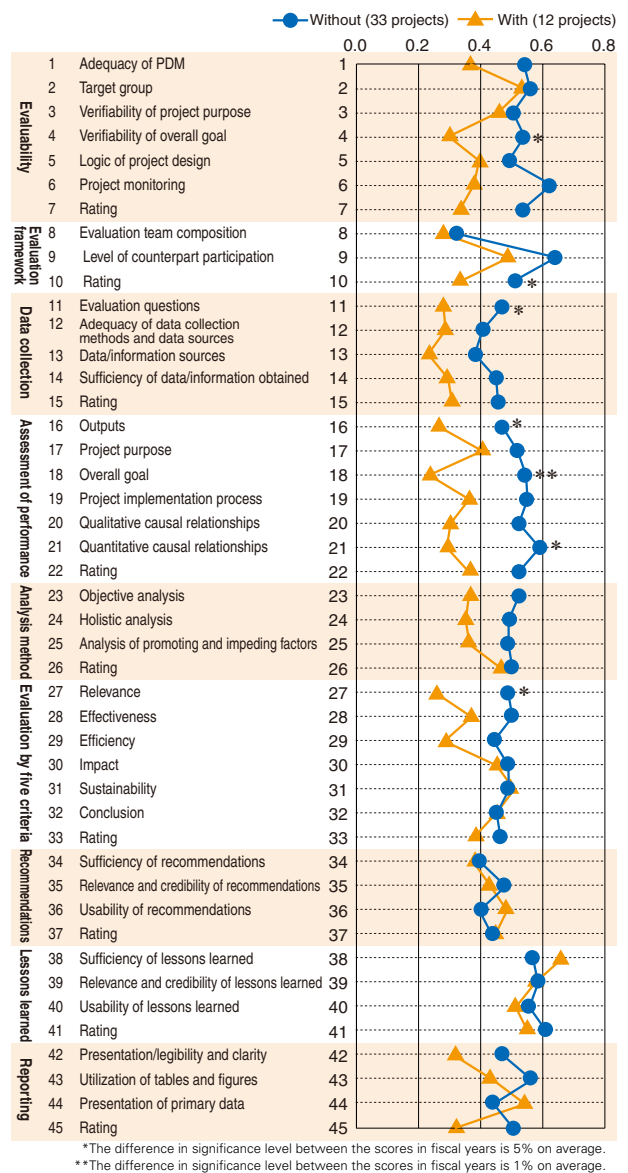
That first batch of projects in which ex-ante evaluations were introduced in 2001 will soon be terminated. Among the secondary evaluations, eight projects in fiscal 2004 and four projects in fiscal 2005 underwent ex-ante examination as such.

There is an assumption that whether or not a project is launched upon the approval of the results of ex-ante evaluation may have considerable influence on the quality of project and the quality of the reports. From the standpoint of consistency of project evaluation from ex-ante to terminal stages, ex-ante evaluation is supposed to set measurable goals at each stage in advance with the same viewpoints as those set at the terminal stage, examine thoroughly the causal relationship between the project purpose and overall goal, collect information on numerical indicators at the initial conditions, and present justified estimates about the change in indicators after the completion of the project. If a project is designed well in the clear causal relationship of structure as well as necessary indicators, it should proceed successfully. If an evaluation report is compiled in accordance with the changes in each indicator, the evaluation report will be convincing and of high quality.

The secondary evaluation this year attempted to compare the average scores between the projects with and without ex-ante evaluations, in addition to the conventional analyses. However, the results of analysis of fiscal 2006 exhibited no significant difference statistically in the two groups for all the evaluation criteria.

Figure 4-10 indicates the degree of variance (standard deviation) of the scores for the projects with and without ex-ante evaluations. The graph clearly shows that the variance of the scores generally tends to be smaller for the group with ex-ante evaluation than those without concerned evaluation. A statistically significant difference is found in the criteria of “evaluability: verifiability of overall goal,” “rating of evaluation framework,” “data collection: evaluation questions,” “assessment of performance: outputs,” “assessment of performance: overall goal,” “assessment of performance: quantitative causal relationship,” and “five evaluation criteria: relevance”; and the variance of the scores for the ex-ante evaluated group is smaller than those for the non-ex-ante evaluated group. In other words, the quality of terminal evaluations is homogenized within the ex-ante evaluated group. Even though the number of projects with ex-ante evaluations is limited, this suggests the possibility that ex-ante evaluation may make terminal evaluation easier. However, considering the small number of target projects and that ex-ante evaluations themselves have gone through improvements and refinements several times since its introduction, re-analysis is warranted for fiscal 2007 to

Figure 4-10 Implementation of Ex-ante Evaluation (Standard Deviation of Scores)



confirm these points.

(7) Summary of the Quality of Primary Evaluation

The overall quality of terminal evaluations attained a certain level; however, the score for “evaluation framework” is relatively lower than other evaluation criteria. In the criterion, the scores for the viewpoints of “composition of evaluation team” and “level of counterpart participation” are both low. An evaluation team is often comprised of a JICA staff member, an evaluation analyst, and one evaluation member. Naturally, when a project stakeholder is included in the evaluation team, it is difficult to judge whether neutrality and fairness are ensured, while specialization is guaranteed. Regarding the level of counterpart participation, it may be because the participation of counterparts at the time of evaluation study is vague and the report does not describe the specialization and neutrality of evaluators of the partner country and

implementation method of evaluations.

The evaluation sheet for fiscal 2006 was revised so as to incorporate the relationship between the viewpoints of “evaluability” and those of “assessment of performance.” Among the viewpoints of “evaluability,” the scores for “verifiability of overall goal” are low and among the viewpoints of “assessment of performance” the scores for “assessment of performance: overall goals” are relatively lower than the other viewpoints. From these results, it is assumed that primary evaluation on assessment of performance may be influenced if the indicators of the overall goal are not clearly identified in the PDM and if the causal relationship between the project purpose and overall goal are vague. Therefore, assessment of performance will be made easier when the indicators for overall goal are appropriately set and when the causal relationships between project purpose and overall goal are clearly defined.

With respect to “efficiency” in the criterion of “evaluation,” many evaluations assessed efficiency from the aspect of the implementation process alone, such as the timing of dispatch of experts and provision of equipment and materials. Only a few evaluations assessed and described efficiency from the aspect of cost efficiency by conducting cost-benefit comparison with other similar projects. It is necessary to conduct primary evaluation with due consideration given to cost effectiveness for the sake of accountability to the people.

The quality of terminal evaluation improves every year, and the secondary evaluation of last year showed that the quality of project evaluations in fiscal 2004 has improved compared to that of fiscal 2003. Such a conclusion is the same as that of the last fiscal year, but it is more reliable since the number of studied evaluations is larger than that in the last fiscal year. Therefore, it has been confirmed that the quality of terminal evaluations in fiscal 2004 has improved.

JICA has adopted “overseas office’s direct project management” on a trial basis since 2004 by endowing the authority to conduct the series of operations for projects. Overseas offices have conducted terminal evaluations on these projects. When terminal evaluations conducted by them are compared to those by the headquarters, the quality of the former tends to be lower. The possible reasons for the low scores include weak logical understanding of project evaluations, insufficient assessment of performance and analyses, and the fact that evaluations were not conducted in line with JICA’s Evaluation Guidelines. The number of projects under overseas office’s direct management is expected to rise in the future, and the guidance for evaluation method, thorough adoption of evaluation standards, and maintenance of the quality of evaluations will have to be addressed.

The frequency of the use of tables and figures is rated low as shown in the criterion of “reporting.” It is thus necessary to use tables and figures and describe them in a clear and understandable manner with the awareness that the reports are readable to the public, let alone project stakeholders. Some evaluations devoted many pages to descriptions of the process of project implemen-

tation, while providing insufficient space for analysis of how the results were attained. It is still desirable to present evaluation results in a convincing way even if the project results turned out to be worse than expected.

(8) Improving Primary Evaluation

As observed above, secondary evaluations were conducted and some valuable comments for increasing the quality of primary evaluation were obtained by the secondary evaluators. Some essential points obtained from these analyses are summarized below.

Some of these points were listed in the secondary evaluation report of fiscal 2005. Since a project takes a few years to be completed, a simple execution of recommendations of fiscal 2005 would not instantly bring about a direct change, but quality would gradually be improved by accumulating such efforts.

Similar factors seen in fiscal 2005 are also identified in fiscal 2006. The following points, though some may overlap, should be key factors for the betterment of the quality of primary evaluation.

1) Timing of Terminal Evaluation

Most terminal evaluations are conducted anywhere from a few months to six months before the termination of a project, in view of the project improvement based on the evaluation results. In the last six months or so of a project various activities are finalized. The current timing of terminal evaluation may hinder the accurate measurement of project achievements. Therefore, it is essential to consider the timing of evaluation, depending on the characteristics of each project and the timing of the expected onset of effects. In the case that the evaluation team concludes in its recommendations that the target will be achieved by the time of the completion of the project, though it will not have been achieved by the time of terminal evaluation, it is desirable to disclose the final results as a response to the recommendation of the evaluation.

2) Composition of Evaluation Team

Considering the quality of evaluations, it is necessary to increase the participation of partner countries in evaluations. Also it is desirable for the report to include descriptions about the evaluators: his/her involvement in the concerned project and/or evaluation method.

3) Data Collection

Sufficient data collection enables the accurate measurement of the achievement of goals. However, in some reports, data collection was insufficient, which lead to an insufficient assessment of performance, resulting in the low quality of evaluation.

Data collection should be in line with PDM, but in the case where data cannot be obtained by the method described in PDM, it is necessary to try to collect data by alternative means. The sources (e.g. questionnaires and interviews) were quite limited, in some cases, to counterparts inside of the implementing organiza-

tion and trainees. It is necessary to expand the scope of data sources from the policy making level to beneficiaries and surrounding communities for confirmation of relevance and process.

As for efficiency, some evaluations did not sufficiently analyze cost efficiency. That is not to say that a vast amount of funds can be spent to achieve the project purpose. Nonetheless, it is necessary to collect data required to evaluate cost efficiency by comparing them with similar projects.

4) Understanding of Important External Assumptions

It is necessary to enhance the assessment of performance regarding the items listed as external factors. When analyzing the effectiveness of projects, it is not enough to confirm only achievement. In order to verify how much the project contributed to the achievement of the goals, it is essential to understand firmly various internal and external factors that may greatly affect the outcomes of a project.

5) Partnership among Projects

Some projects are follow-up projects or a part of grant aid. In such cases, it is not always clear whether the achievements are the results of the project itself or those of related projects. Some reports did not mention the related project or its effects, and thus it is necessary to convince the readers by referring to the related information.

6) Objective Analysis

Objectivity cannot be reached solely through the results of quantitative analysis. Some reports only listed positive results and did not describe how it was achieved. Even in qualitative analysis, it is important to describe the reason for the results and the process by which they were attained, and objectivity can be achieved by providing the logical reasons for such conclusions.

7) Viewpoint for Assessment of Impact

Since terminal evaluation is conducted several months prior to the actual termination of a project, it seems to be a viable approach to evaluate the degree of achievement of the goal/purpose, relevance, and efficiency, and to compile recommendations about sustainability through the assessment of the implementation system. However, this is not the case with the impact, which is the prediction as to how much impact (positive/negative) is brought about in the future. In some cases it is questionable whether the impact will actually happen by the time of the termination of the project. It is necessary to discover an impact, no matter how small it may be, to present as the basis for the prediction and to increase the credibility, instead of presenting wishful thinking.

8) Report Writing

There are some reports without a PDM or an evaluation grid. It is necessary to present guidelines for writing a report. Even if

the project period is less than a year, the report needs to be in line with the guidelines. As it is expected that more and more will be managed and evaluated by overseas offices in the future, more detailed guidelines are necessary to maintain the quality of evaluation.

A report should start with a summary, followed by the main text describing details and raw data attached at the end. The main text is more understandable if it follows the PDM procedures, starting with goals and proceeding with activities carried out to achieve the goals.

Tables and figures illustrated concisely and clearly would be a good tool for readers. However, not a few reports went too far in generalization by simply stating in the text, for example, "interests of students in science increased," giving no basis for such a statement, and readers need to look into the evaluation grid to find the details (in some cases, the details are still unclear). It is understandable to readers if the report contains some detailed and logical descriptions in the text by, for example, presenting concise data for the basis of important items.

Other cases provided the results of analysis on promoting and impeding factors in the summary, although the main text mentioned no such thing. It is necessary to ensure consistency between the main text and the summary.

Some reports contained a large number of supplementary documents as attachments, such as the list of participants in the seminar, while the main text itself was short. There were many attachments that were not referred to in the main text and seemed to be unnecessary. A line-up of facts and lists that seem to be unrelated to the evaluation or analysis should be excluded from the report, and instead should be included only in a CD-ROM.

In sum, what is most important about the report writing is that it be easy to read, especially for the general public. In areas such as basic research and medicine, in particular, when an achievement is made in an experiment, it becomes more understandable to non-experts if an additional explanation is given as to how much impact the concerned achievement has in a broader perspective. For example, success in the cultivation of a microorganism is the first step toward the development of a drug for early detection of a disease. Furthermore, it becomes more understandable if the report states how long it would take for the development of drug from the cultivated microorganism.

9) Role of Terminal Evaluation and Secondary Evaluation

The purpose of evaluation is first to confirm the facts and make a judgment about the success or failure of each fact that follows. In the case where only a judgment is mentioned with insufficient confirmation of the facts, the issue is whether or not the reader will believe the content. If the facts are accurately confirmed, the reader will be able to make a judgment different from that of the report about the facts.

First of all, it is important in terminal evaluation to accurately understand the achievements of the project purpose and the process that leads to the achievements. After confirming a fact,

judgment should be made about the fact, followed by a confirmation of the implementing system of the partner country, a discussion of future support systems, and a compilation of recommendations and lessons.

When conducting the secondary evaluation, if facts have been sufficiently established at the time of terminal evaluation, the secondary evaluators are able to make judgments based on the facts of the terminal evaluation report, thus allowing them to compare their judgments with those of the primary evaluators.

1-3 Project Evaluation by Secondary Evaluators Based on Terminal Evaluation Reports

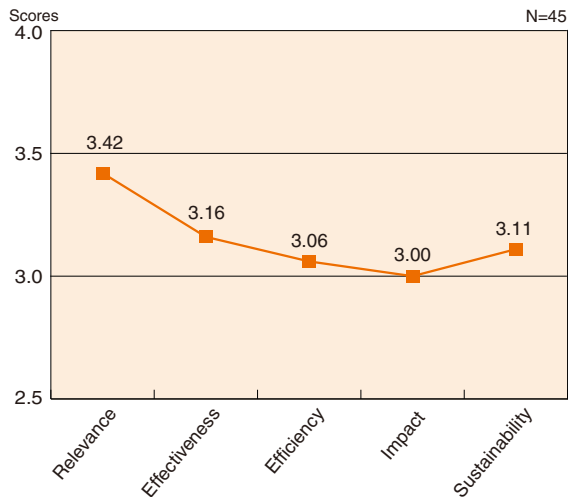
(1) Summary of the Secondary Evaluation of Projects

We conducted the secondary evaluation on 28 projects evaluated in fiscal 2004 and 17 projects in fiscal 2005 using the terminal evaluation reports from the perspective of the five DAC Evaluation Criteria. Figure 4-11 shows the result of the project evaluation gleaned from the reports by the secondary evaluators.

All the average scores for the projects are in the 3-point range, being above the level of “average.” Among the five evaluation criteria, the average scores for “relevance” was the highest at 3.4 points, while the average scores for “efficiency” and “impact” are relatively low at 3.1 points or lower.

Figure 4-12 illustrates the distribution of scores for the project evaluation. All of the scores for “relevance” are over 2.5 points, including some at 4.0 points and higher. Most of the scores are clustered between 3.5 and 3.99 points with a small variance and generally in the higher range. The scores for “effectiveness” and “sustainability” are evenly clustered between 2.0 and 3.99 with some given 4.0 points or higher. No more than 4.0 points are given to “efficiency,” and although some fall in the range of 3.5 and 3.99, most scores fall in the range of 2.5 and 3.49, showing relatively low ratings. As for the scores for “impact,” some are in the 1-point range and none are 4.0 points or higher, showing low ratings.

Figure 4-11 Project Evaluation by Secondary Evaluators



(2) Project Evaluation from Viewpoints for Each Criterion

We conducted secondary evaluation from various viewpoints for the five evaluation criteria based on the information obtained

Figure 4-12 Distribution of Evaluation Scores for Projects by Secondary Evaluators

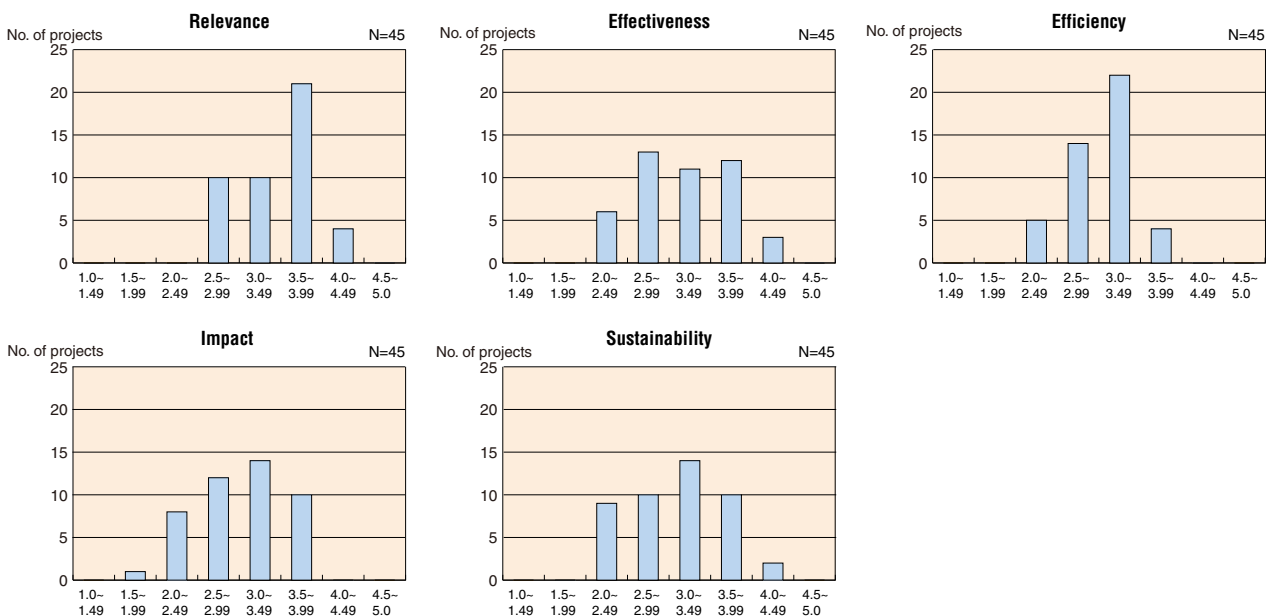
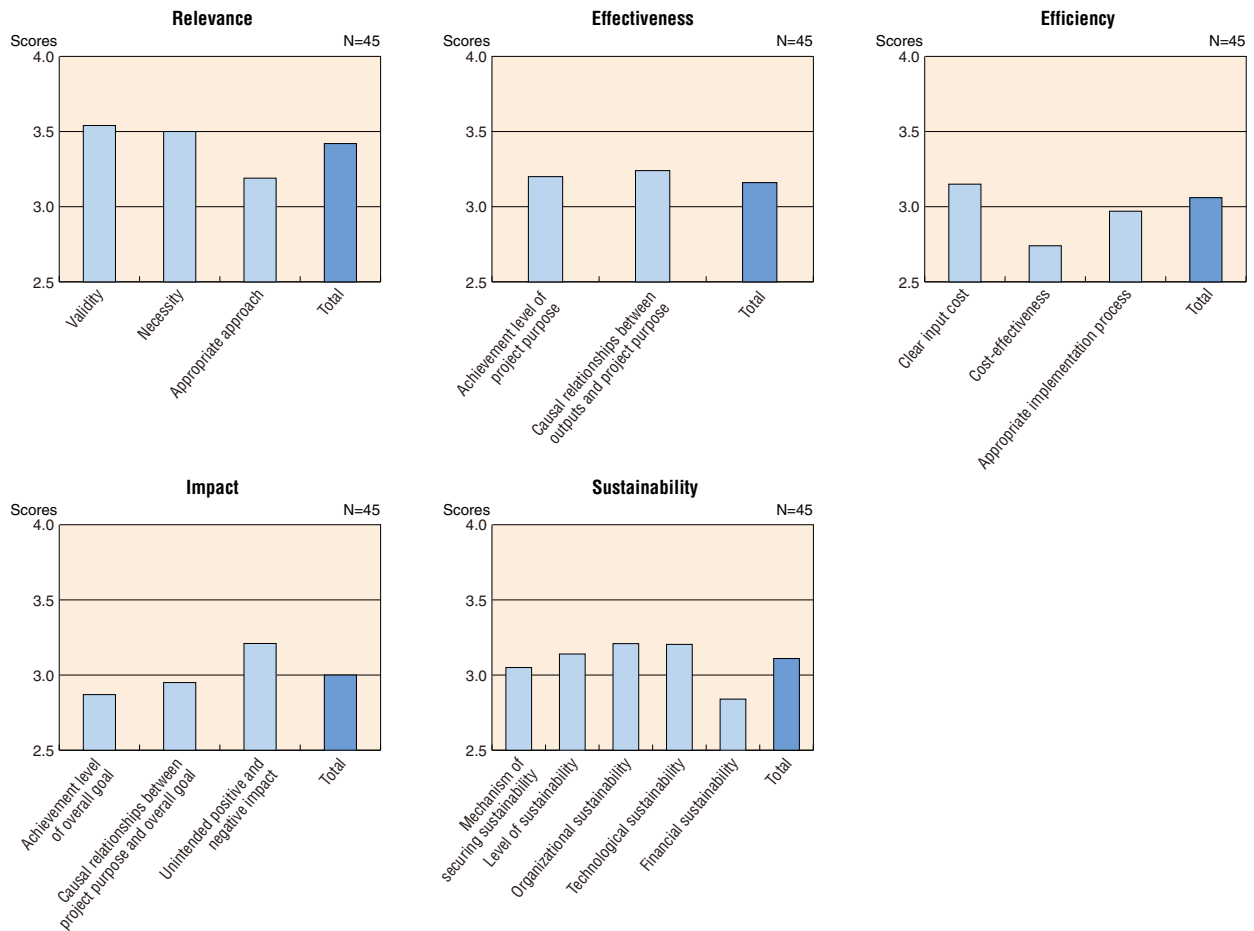


Figure 4-13 Project Evaluation by Secondary Evaluators: Viewpoints (Average Scores)



from the reports. Figure 4-13 shows the average scores for the viewpoints.

1) Relevance

The average scores for all the viewpoints of “relevance” are high. Among the three viewpoints, the average scores for “validity” and “necessity” of project implementation are high with more than 3.5 points. The average score for “appropriate approach” as to whether the approach was appropriate and effective for achieving project purposes is 3.2 points, relatively lower than those for other viewpoints.

A project that implemented drug control targeting five countries were evaluated high not only on “validity” and “necessity” but also on “appropriate approach” because of the good design of the project implemented as regional cooperation. On the other hand, the projects that aimed to improve cattle hygiene or develop rural infrastructure were evaluated somewhat favorably on “validity” of the project purpose and “necessity,” while they were rated low on “appropriate approach” based on the conclusion that the approach of enhancing the research institute or technical center to achieve the project purposes was not appropriate.

2) Effectiveness

In the criterion of “effectiveness,” “achievement level of

project purpose” and “causal relationships between outputs and projects purpose” received a relatively high average score of around 3.2 points.

Those projects in which indicators for the achievement of purposes were all achieved and the outputs led to outcomes were evaluated high on “effectiveness.” However, a project was evaluated low on “effectiveness” if the external factor of delayed dispatch of experts due to late visa issuance failed to achieve the outputs, resulting in the project purposes not being achieved.

3) Efficiency

In the criterion of “efficiency,” the average score for “clear input cost,” a viewpoint to measure if the input cost, for example, for the purchase of equipment and dispatch of experts was made clear, is relatively high with 3.2 points. However, the score for “cost-effectiveness,” to see if efforts to achieve more effects with lower costs were made, is low with 2.8 points, which is the lowest rating of all the viewpoints under all the evaluation criteria.

The projects that were evaluated high on “efficiency” made efforts to reduce costs through local procurement of equipment, maintain consistency in technical transfer through training by short-term experts to their counterparts at the agencies where they work, and save costs through utilizing the equipment and facilities on the side of the partner countries. On the other hand,

some projects were evaluated low, because “although a very expensive compost plant was constructed, it is not utilized well” and that “dispatch of experts and counterparts was delayed.”

4) Impact

Among the three viewpoints under the criterion of “impact,” the average score for “achievement level of overall goal” to see if the planned effects were achieved or likely to be achieved through the achievement of project purposes was the lowest at 2.9 points. On the other hand, average score for “unintended positive and negative impact” to see an emergence of unintended impact such as economic impact on policies, target society and beneficiaries, and gender equality was the highest at 3.2 points.

One of the projects that was evaluated high on “impact” achieved its project purpose of establishing a housing evaluation system, producing impacts such as increased supply and demand for good housing, and resulted in unintended impacts; for example, a movement to establish a technical standard for housing parts that were introduced by the implementation of the project emerged and the local government started to establish an implementing agency for the national implementation of the system.

5) Sustainability

Most viewpoints were more than 3.0 points on the average. Among the three viewpoints in relation to the establishment of a mechanism for securing sustainability, the average scores for “organizational sustainability,” which concerns the organizational capability of securing sustainability, and “technological sustainability,” which measures whether the techniques and skills obtained through a project are maintained or expanded, are relatively high at 3.2 points. On the other hand, the average score for “financial sustainability,” which questions if any measures were taken to secure budgets, is 2.8 points, which is the lowest among

all the viewpoints under “sustainability.”

In a project aimed at improving the regional medical care system, most counterparts who received technical transfer continue to be involved in the activities. The project was evaluated high on “sustainability” because transferred technology can be maintained and managed, the implementation organization has sufficient capabilities to maintain and develop the effects of the project, and reliable political and financial support were obtained.

(3) Analysis by Viewpoint for Relevance

When looking at project implementation chronologically, a project is started with the relevance of the project examined. Let’s take a look at the connection between the relevance evaluated at the beginning of a project and the subsequent implementation process and outcomes.

Table 4-7 illustrates the correlation between the scores for the viewpoints in the criterion of “relevance” and those for viewpoints and evaluation criteria of “effectiveness,” “efficiency,” “impact,” and “sustainability.” It is clear that viewpoints for relevance have a high level of correlation with the scores for evaluation criteria of “effectiveness,” “efficiency,” and “sustainability,” with the highest level of correlation with “impact.” “Relevance” is evaluated from the three viewpoints of “validity,” “necessity,” and “appropriate approach.” Of these, “appropriate approach” has the highest correlation to other evaluation criteria in the ratings, especially to “achievement level of project purpose” and “causal relationships between outputs and project purpose” for “effectiveness,” “achievement level of overall goal” and “causal relationships between project purpose and overall goal” for “impact,” and “level of sustainability” and “organizational sustainability” for “sustainability.”

Therefore, it can be concluded that “appropriate approach” for “relevance” in the project is highly related to the emergence of

Table 4-7 Correlation between Relevance and Other Four Criteria

Evaluation Criteria/Viewpoints		Relevance			
		Validity	Necessity	Appropriate Approach	Overall Rating
Effectiveness	Achievement level of project purpose	0.330 *	0.424 **	0.647 **	0.549 **
	Causal relationships between outputs and project purpose	0.545 **	0.624 **	0.674 **	0.626 **
	Overall rating	0.471 **	0.525 **	0.725 **	0.632 **
Efficiency	Clear input cost	0.392 **	0.455 **	0.448 **	0.473 **
	Cost-effectiveness	0.334 **	0.495 **	0.480 **	0.522 **
	Appropriate Implementation process	0.390 **	0.499 **	0.589 **	0.613 **
	Overall rating	0.405 **	0.497 **	0.627 **	0.621 **
Impact	Achievement level of overall goal	0.517 **	0.544 **	0.697 **	0.696 **
	Causal relationships between project purpose and overall goal	0.580 **	0.538 **	0.645 **	0.647 **
	Unintended positive and negative impact	0.429 **	0.447 **	0.580 **	0.625 **
	Overall rating	0.586 **	0.632 **	0.738 **	0.786 **
Sustainability	Mechanism of securing sustainability	0.475 **	0.559 **	0.599 **	0.624 **
	Level of sustainability	0.532 **	0.588 **	0.671 **	0.666 **
	Organizational sustainability	0.454 **	0.518 **	0.622 **	0.632 **
	Technological sustainability	0.262	0.402 **	0.464 **	0.361 *
	Financial sustainability	0.426 **	0.438 **	0.488 **	0.481 **
	Overall rating	0.449 **	0.544 **	0.558 **	0.604 **

*Correlation is shown at 5% of significance level.

**Correlation is shown at 1% of significance level.

project outcomes and sustainability and also it is important to select appropriate means and methods in implementing a project.

(4) Project Evaluation by Sector

Projects are implemented across a variety of sectors. The evaluation target can be categorized: 24 projects in the sector of social development, seven in agricultural development, six in forest and natural environment, six in health and medical care, and two in mining and industrial development. The projects in social development include human resources development, rural development, and pollution prevention. Agricultural development includes agriculture promotion, rural environment conservation, and enhancement of a veterinary research institute. The projects in forest and natural environment include forest fire prevention, water management improvement, and afforestation techniques for degraded area. The projects in health and medical care include regional medical care improvement, maternal and child health, and medicine management. The projects in mining and industrial development include energy control.

Figure 4-14 and Table 4-8 illustrate the project evaluation by sector. Although the number of projects evaluated varies from sector to sector, the evaluation results exhibit differences by sector. All the average scores for each evaluation criterion in the sectors of social development, health and medical care, and mining and industrial development, are more than 3.0 points, achieving the level of “average” or higher. A similar evaluation tendency is observed among social development, health and medical care, and mining and industrial development, with relatively lower

scores given for “effectiveness,” “efficiency,” and “impact” as compared to “relevance” and “sustainability.”

As for the projects in health and medical care, Project for Development of Trauma Center Complex in Thailand, was a technical cooperation project aimed at establishing a model for trauma prevention in line with the actual conditions of local cities in order to enhance the care and prevention of trauma caused by traffic accidents. The regional emergency system was established earlier than the national system, resulting in a lower death rate by traffic accident. As impacts from the project, hospitals in areas other than the target area started to refer to the project activities and emergency life guards who were trained and recruited through the project obtained national accreditation. The average score for “relevance” of this project is 4.1 points; for “effectiveness,” “efficiency,” and “impact” they are all 3.8 points or higher, and for “sustainability” the score is 3.5, resulting in a very high rating.

In contrast, in the sectors of agricultural development and forest and natural environment, the average scores for many evaluation criteria show a similar tendency, falling in the range from 2.0 to less than 3.0. In those sectors, the average scores for “impact” and “sustainability” are lower than those for “relevance,” “effectiveness,” and “efficiency.” As clearly shown in Table 4-8, the average scores in the sectors of rural development and forest and natural environment are lower than those in the sectors of social development and health and medical care, showing significant difference. The tendency of average scores being generally lower in the sectors of agricultural development and forest and natural environment than those in the other sectors was also observed in the results of the secondary evaluation in fiscal 2005.

The evaluation on six projects in the forest and natural environment sector shows a variance. The Forest Fire Prevention Management Project (Phase 2) in Indonesia was a technical cooperation project that set the project purpose as conducting forest fire prevention management activities in order to protect national parks from forest fires. In the project, the comprehensive forest fire prevention management model was about to be completed at the target national park. The total average score for the five evaluation criteria for this project is 3.4 points, with all the average scores for every criterion at 3.2 points or higher, including the average scores for “effectiveness” and “sustainability” at 3.7, resulting in a high rating. On the other hand, total average scores of four projects in the sector are 2.8 points or lower. The lowest score in overall evaluation was given to Technology Development for Revegetation and Utilization of Degraded Areas in the Semi-arid Region of the Northeastern Brazil, whose project purpose was the recovery of degraded area and prevention of desertification. Due to a significant delay in dispatching experts, the project purpose has not been achieved and it is now difficult to forecast the achievement of the overall goal. The average scores for all the evaluation criteria for this project are in the 2-point range.

Figure 4-14 Evaluation by Sector (Average)

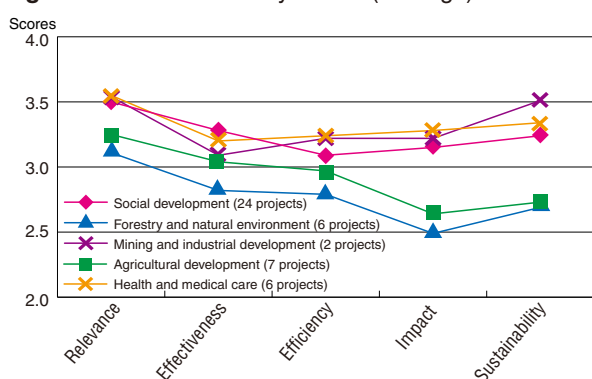
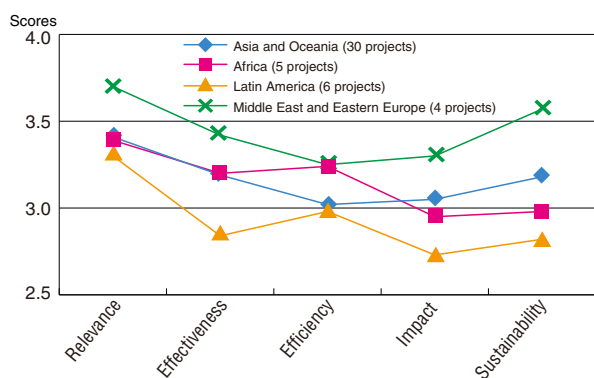


Table 4-8 Evaluation by Sector (Average)

	Social Development	Health and Medical Care	Mining and Industrial Development	Agricultural Development	Forestry and Natural Environment	Difference in Average Scores
Relevance	3.50	3.55	3.54	3.25	3.11	
Effectiveness	3.28	3.20	3.09	3.04	2.82	
Efficiency	3.09	3.24	3.22	2.97	2.79	
Impact	3.15	3.28	3.22	2.64	2.49	**
Sustainability	3.24	3.34	3.51	2.73	2.69	*
Total	24	6	2	7	6	

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

**The difference in significance level among sectors is 1% on average.

Figure 4-15 Evaluation by Region (Average)**Table 4-9** Evaluation by Region (Average)

	Asia and Oceania	Africa	Latin America	Middle East and Eastern Europe	Difference in Average Scores
Relevance	3.41	3.39	3.30	3.70	
Effectiveness	3.19	3.20	2.84	3.42	
Efficiency	3.02	3.24	2.98	3.25	
Impact	3.03	2.95	2.73	3.30	
Sustainability	3.13	2.98	2.82	3.57	*
Total	30	5	6	4	

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

(5) Project Evaluation by Region

Projects are widely implemented across regions. The number of projects subject to the secondary evaluation in fiscal 2006 is 30 in Asia and Oceania, six in Latin America, five in Africa, two in the Middle East, and two in Eastern Europe. As the number of projects in the Middle East and Eastern Europe is small and the mean scores for the evaluation criteria show a similar tendency, these two regions were merged into one group for analysis purposes. Figure 4-15 and Table 4-9 indicate the result of evaluation by region.

As for the average scores by region, “sustainability” was rated higher in the Middle East and Eastern Europe than in Latin America, with a significant difference statistically; however, no significant differences were observed among the regions for the other criteria. However, as evidenced in Figure 4-15, Asia and Oceania, the Middle East, and Eastern Europe show a similar tendency, where “relevance” and “sustainability” were rated relatively high while “efficiency” was rated the lowest of the evaluation criteria. Africa and Latin America also show a similar tendency for the evaluation criteria, where “relevance” and “efficiency” were rated relatively high and “impact” was rated the lowest.

In fiscal 2005, the studied evaluations of the Middle East received low marks for every criterion compared to the other regions, while relatively good marks by Latin America. In fiscal 2006, however, Middle East and Eastern Europe received relatively high ratings as shown in Figure 4-15 as opposed to the results of Latin America.

The projects in Latin America were rated relatively high or

Table 4-10 Distribution of the Target Projects by Region and Sector

Sector \ Region	Asia and Oceania	Africa	Latin America	Middle East and Eastern Europe	Total
Social development	18	2	2	2	24
Agricultural development	4	1	2	0	7
Forestry and natural environment	4	1	1	0	6
Health and medical care	3	1	1	1	6
Mining and industrial development	1	0	0	1	2
Total	30	5	6	4	45

low, showing a wider variance among the projects. For example, Development of Method of Research and Education in Electric Field in Mexico was a technical cooperation project aimed at enhancing the capacity of instructors for preparing teaching materials at vocational schools using new electric technologies. The impacts could be observed, for example, where organizations other than counterparts became interested in many applicable teaching materials developed with the advanced technology, and thanks to easier communications with other countries, they are planning the formulation of teaching materials in cooperation with the other countries. This project was rated relatively high, with average scores of 3.2 points or higher. In contrast, in the Improvement of the Small-medium Dairy Farm Management Project in Paraguay, aimed at establishing a dairy management model suitable for the management improvement of small and medium-size farms through dairy farming, the project activities as a target deviated from the original project purpose because the causal relationships between the indicators of the project purpose and the level of achievement were not appropriate. Though the average score for “relevance” of this project are in the 3-point range, the average scores for all other criteria are in the 2-point range. Among the other project studied in Latin America, the above-mentioned Technology Development for Revegetation and Utilization of Degraded Areas in the Semi-arid Region of Northeastern Brazil is included.

Table 4-10 lists the projects based on regions and sectors of the projects. As shown in the table, half of all the projects implemented in Latin America are in the agricultural development and forest and natural environment sectors. The projects in those sectors were rated lower than projects in other sectors as shown in the project evaluation by sector. This likely led to the low overall rating for the Latin America region.

(6) Project Evaluation by Year

Figure 4-16 and Table 4-11 show the changes in average scores for project evaluation by year. As shown in Figure 4-16, all the average scores for all evaluation criteria of projects in fiscal 2004 and 2005 are higher than those in fiscal 2003. Table 4-11 indicates the results of the statistical analysis. There is a statisti-

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

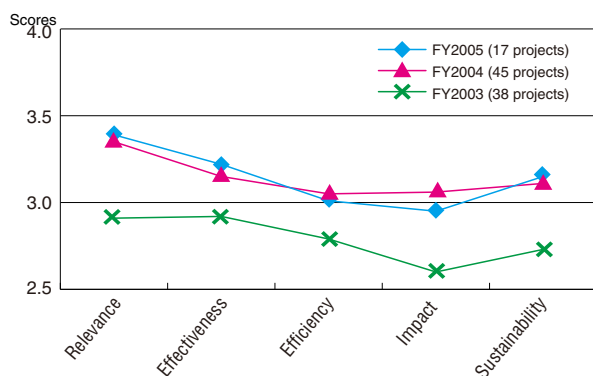


Table 4-11 Year-to-Year Changes of Project Evaluation by Secondary Evaluators (2003-2005)

Evaluation Criteria	Average Score			Difference in Average Scores		
	FY2003 (A)	FY2004 (B)	FY2005 (C)	(B)-(A)	(C)-(A)	(C)-(B)
Relevance	2.91	3.35	3.39	0.44**	0.48**	0.04
Effectiveness	2.92	3.15	3.22	0.23*	0.30*	0.08
Efficiency	2.79	3.05	3.01	0.25**	0.21	-0.04
Impact	2.60	3.06	2.95	0.46**	0.35*	-0.11
Sustainability	2.73	3.11	3.15	0.38**	0.43**	0.05

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

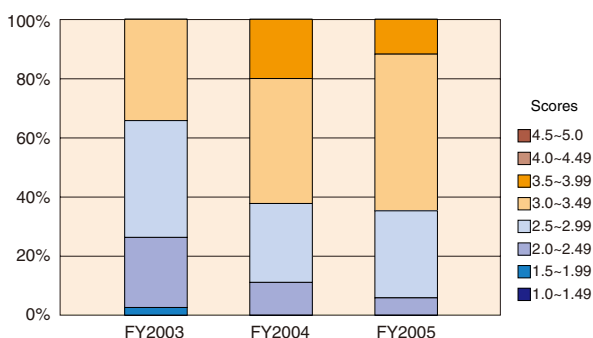
** The difference in significance level between the scores of fiscal years is 1% on average.

cally significant difference in average scores for all evaluation criteria between the projects of 2003 and 2004, and the scores of 2004 are rated higher. Comparing the projects in 2003 to those in 2005, the latter are rated higher for all evaluation criteria, and a statistically significant difference can be observed in the scores for “relevance,” “effectiveness,” “impact,” and “sustainability.” However, there are no significant differences between 2004 and 2005, which received similar evaluations for all criteria.

In Figure 4-17, the distribution of total scores or the average of all the evaluation scores is projected. As shown on the bar charts, though in 2003, there were projects whose total scores remain in the range from 1.0 to less than 2.0, there are no such projects in fiscal 2004 and 2005. The number of projects that were rated less than 2.5 points is decreasing in fiscal 2004 and 2005, and the ones with more than 3.5 points increased.

The results of last fiscal year’s analysis showed that projects

Figure 4-17 Distribution of Total Scores (Average Score)



in fiscal 2004 were rated higher by secondary evaluators than those in fiscal 2003. Comparing the project evaluation between fiscal 2003 and fiscal 2004, similar to the results obtained last fiscal year, projects implemented in fiscal 2004 were rated higher than those in fiscal 2003. The results of this fiscal year show that the scores for the projects in fiscal 2004 and fiscal 2005 are significantly higher than those in fiscal 2003 from a statistical point of view. Last fiscal year’s analysis targeted 17 projects in fiscal 2004, while this fiscal year targeted 45 in fiscal 2004, providing steadier results. Although the evaluations are based on the reports, from the above results, it was confirmed that projects in fiscal 2004 were rated higher than those in fiscal 2003.

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

Following the trend of decentralization, JICA has adopted “direct project management by overseas offices” at eight overseas offices on a trial basis since October 2004. Under this system, the overseas offices are entrusted with the authority to carry out the series of operations from project formulation to implementation, and project evaluation. Since April 2005, the number of supervising offices had expanded to 30. Comparing projects managed by the headquarters and those by overseas offices, though there were differences in the evaluation quality of terminal evaluation reports as mentioned in Section 1-2-(5), no differences were observed in the results of the project evaluation gleaned from the reports by secondary evaluators.

In addition, since the projects on which ex-ante evaluations had been introduced will soon be terminated, among the projects studied this year, 12 projects underwent ex-ante examination. No differences between the projects with and those without ex-ante evaluation were observed in the results of the project evaluation by secondary evaluators.

However, the numbers of terminal evaluations is quite limited for direct management projects by overseas offices and projects with ex-ante evaluation, so the analysis result of this fiscal year is far from being definite. In the future, it is expected that the greater the number of terminal evaluations for projects managed by overseas offices and those that underwent ex-ante evaluation, the more stable will be the results obtained.

(8) Macro Factors Affecting Project Outcomes

1) Analysis by Correlation

In the previous section, we made the comparative analysis on evaluation results for various outcome indicators of projects in terms of aid sector and target region. With this analysis, it is clear that the outcomes of projects (results of five evaluation criteria) differ depending on the sector and region. However, they are not the only factors that can affect the outcomes of projects. For example, “project scale” can be such an influential factor. It might be assumed that the larger the project scale, the bigger the impact, which makes it easier to generate outcomes; or, it can be

assumed in an opposite way, this makes it more difficult to maintain project outcomes. The cooperation period can also be a potential factor: assumed that the longer the cooperation period, the easier it is to obtain outcomes. Or, on the contrary, a long cooperation period will make it difficult to facilitate outcomes, which may lead to fewer effects.

The conditions of target countries also affect the facilitation of outcomes. For example, since various physical, technological, and organizational conditions are required to generate project outcomes, it can be assumed that factors such as overall financial strength, organizational management capacity, level of governance and general intellectual standard of the partner country impact the outcomes. It is also assumed that the level of those impacts varies by outcome index. Finally, considering the relations between sector and region and these macro factors, it is also important to decide what really impacts the project outcomes.

Therefore, we will focus on not only “project scale” and “cooperation period,” but also “GDP/cap” as an indicator of financial strengths of the partner country, “human development index (HDI)” for the organizational management capacity and level of governance, and “education index” which indicates the general intellectual standard, in order to see the relations with projects. This analysis targeted 60 projects that second evaluators in fiscal 2006 directly assessed (45 newly added in fiscal 2006 and 15 projects as overlaps evaluated in fiscal 2005). As for “project scale” and “cooperation period,” the data/information was obtained from the terminal evaluation reports. The values for “GDP/cap,” “HDI,” and “education index” were taken from the

Human Development Report (UNDP, 2006). “Education index” refers to the synthesized index of the adult literacy rate and enrollment rates of primary, secondary, and higher education, and “HDI” is the index synthesizing the average mortality at the time of birth, education index, and GDP. Logarithm natural figures were used for “GDP/cap.”

In Table 4-12, the correlations between the scores for the criteria and viewpoints of five criteria evaluation and the above-mentioned factors for 60 projects are indicated. As shown clearly in the table, the scores for “relevance,” “effectiveness,” “efficiency,” and “impact” among the five evaluation criteria are not largely correlated to these factor variants except for a few. In other words, regardless of the conditions of these factors, some projects achieved outcomes while others did not.

On the other hand, when it comes to “sustainability,” both the criterion and its viewpoints show some close correlations with external factors that are not in direct relation to projects such as “GDP/cap,” “HDI,” and “education index.” In particular, the criterion and each of the viewpoints have strong correlations with “GDP/cap” and “HDI.” Among the viewpoints, technological sustainability, organizational sustainability, and financial sustainability show strong correlations. Because of the strong correlations among “GDP/cap,” “HDI,” and “education index,” it is difficult to determine which is the specific factor that affects the facilitation of outcomes. However, it is obvious that factors such as the financial strength, organizational management capacity, governance and general intellectual standard of the partner country affect sustainability in a broad sense. Also, the result of “sustainability” is convincing enough, because the factor usually

Table 4-12 Correlation between Project Evaluation with Five Evaluation Criteria and Various Factors

Evaluation Criteria		Cooperation Period	Project Scale	GDP/cap	HDI	Education Index
Relevance	Validity	0.157	0.043	0.112	0.137	0.147
	Necessity	0.143	-0.010	0.171	0.185	0.198
	Appropriate approach	0.081	-0.031	0.183	0.147	0.078
	Score	0.148	0.044	0.150	0.158	0.123
Effectiveness	Achievement level of project purpose	0.129	0.093	0.025	0.053	0.042
	Causal relationships between outputs and project purpose	0.137	0.121	0.142	0.172	0.161
	Score	0.146	0.105	0.045	0.071	0.053
Efficiency	Clear input cost	0.205	0.190	0.059	0.110	0.274 *
	Cost-effectiveness	-0.005	-0.145	0.091	0.124	0.154
	Appropriateness of Implementation process	0.030	0.111	0.159	0.152	0.080
	Score	0.037	0.117	0.198	0.221 +	0.148
Impact	Achievement level of overall goal	0.148	0.117	0.099	0.142	0.093
	Causal relationships between project purpose and overall goal	0.083	0.110	0.215 +	0.202	0.151
	Unexpected positive and negative impact	0.040	0.055	0.141	0.171	0.193
	Score	0.179	0.099	0.168	0.188	0.125
Sustainability	Mechanism of securing sustainability	0.128	0.097	0.252 +	0.252 +	0.186
	Level of sustainability	0.245 +	0.193	0.231 +	0.262 *	0.211
	Organizational sustainability	0.128	0.183	0.358 **	0.339 **	0.291 +
	Technological sustainability	0.240 +	0.228 +	0.403 **	0.407 **	0.399 **
	Financial sustainability	0.129	0.228 +	0.286 *	0.292 *	0.232 +
	Score	0.066	0.131	0.366 **	0.358 **	0.317 *

*Correlation is shown at 10% of significance level.

**Correlation is shown at 5% of significance level.

**Correlation is shown at 1% of significance level.

Figure 4-18 Top 4 Projects

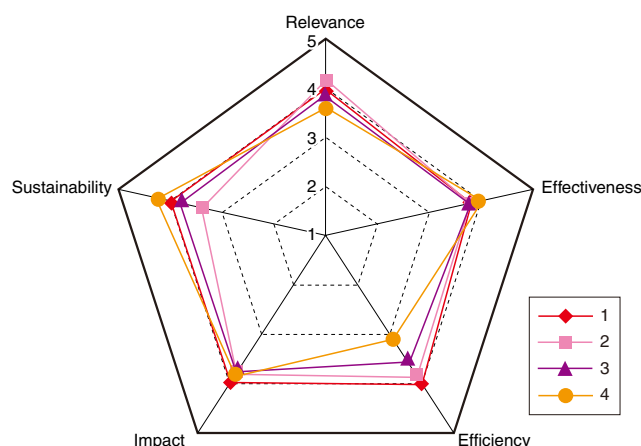
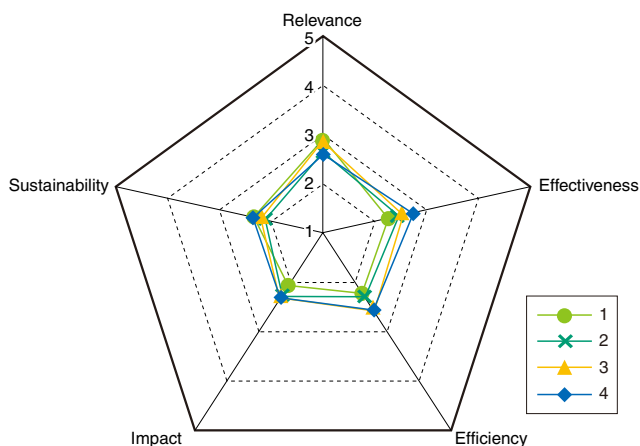


Figure 4-19 Worst 4 Projects



depends on the efforts and systems of the partner country after the project is terminated.

In addition, the fact that one of the viewpoints, “mechanism of securing sustainability,” does not have any strong correlation with the external factors compared to other viewpoints can indicate that mechanism of securing sustainability is incorporated into a project under any conditions and at the same time, it can also be concluded that it is difficult to secure sustainability without the external conditions provided.

2) Necessity of Factor Analysis

Many factors other than these above-mentioned factors can affect the project outcomes. In addition to the socioeconomic variants, the political condition is also an important variant.

Because these factors are strongly correlated to one another, they have to be considered together to determine the effective factor. In order to stabilize the results, a sufficient number of samples (evaluation results) needs to be collected. This analysis is worth a try when sufficient samples are collected in and after fiscal 2007.

(9) Total Project Evaluation by Secondary Evaluators

Based on the aggregates of scores for the five evaluation criteria, which the secondary evaluators gave by judging from the reports (1 to 5 points), the top four projects and worst four projects were selected from 45 projects in fiscal 2004 and fiscal 2005, all of which were subject to this fiscal year’s secondary evaluation. Figure 4-18 and Figure 4-19 indicate the total evaluation

Table 4-13 Differences between Top 4 and Worst 4 (Average Scores)

Evaluation Criteria/Viewpoints		Average Scores		Difference in Average Scores
		Top 4 Projects	Worst 4 Projects	
Relevance	Validity	3.79	3.16	0.62 *
	Necessity	3.81	3.09	0.72 *
	Appropriate approach	3.99	2.44	1.54 **
	Score	3.86	2.73	1.13 **
Effectiveness	Achievement level of project purpose	3.88	2.40	1.48 **
	Causal relationships between outputs and project purpose	3.79	2.88	0.92 **
	Score	3.87	2.51	1.36 **
Efficiency	Clear input cost	3.36	2.61	0.75 *
	Cost-effectiveness	3.07	2.27	0.79
	Appropriateness of Implementation process	3.66	2.46	1.20 **
	Score	3.59	2.37	1.22 **
Impact	Achievement level of overall goal	3.81	2.11	1.70 **
	Causal relationships between project purpose and overall goal	3.75	2.41	1.35 **
	Unexpected positive and negative impact	3.84	2.31	1.53 **
	Score	3.81	2.19	1.62 **
Sustainability	Mechanism of securing sustainability	3.73	2.32	1.41 **
	Level of sustainability	4.01	2.24	1.76 **
	Organizational sustainability	3.90	2.29	1.61 **
	Technological sustainability	3.77	2.59	1.17 **
	Financial sustainability	3.88	2.02	1.86 **
	Score	3.86	2.24	1.62 **

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

scores for those projects in terms of the five evaluation criteria. Table 4-13 compares the average scores for the evaluation criteria and viewpoints for the top four projects and the worst four project group.

As shown in Figure 4-18, Figure 4-19, and Table 4-13, the average overall scores for “relevance,” “effectiveness,” and “sustainability” of the top project group are around 3.9 points or more, while the average score for “efficiency” are relatively low with 3.6 points, showing a variance in ratings. As for the average overall scores for the five evaluation criteria of the worst project group, “impact” and “sustainability” are rated relatively low while “effectiveness” is rated relatively high, although there is a variance in ratings.

Comparing the differences in average scores between the top project group and the worst project group, there is a statistically significant difference for each evaluation criterion, and the first group is rated higher than the second one. Between the two groups, the difference in average scores for “relevance” is small while that for “impact” and “sustainability” is large.

Next, let’s analyze the differences between the “good” projects and the “bad” projects based on each evaluation criterion. Comparing the scores for each viewpoint between the top project group and the worst project group, the average scores for any viewpoint in the criteria for “impact” and “sustainability” is higher in the good project group, and differences in the average scores between both groups are statistically significant. However, as for “relevance,” “effectiveness,” and “efficiency,” a disparity was observed in the differences between both groups in some viewpoints. The difference in average scores for “appropriate approach” in “relevance,” “achievement level of project purpose” in “effectiveness,” and “appropriate implementation process” in “efficiency” was large between the two groups and the average score for every viewpoint in the top four group is higher.

The highly rated projects received high scores for “appropriate approach” in “relevance,” “achievement level of project purpose” in “effectiveness,” “appropriate implementation process” in “efficiency,” and all the viewpoints in “impact” and “sustainability.” Based on these results, it can be assumed that a project is rated high when appropriate means are adopted at the project

implementation, the achievement level of project purpose is high, and an appropriate implementation process is adopted so that spreading effects are observed and lead to a high level of sustainability. In other words, it is indicated that approach and methods of project implementation are important factors for the success of a project.

Table 4-14 and Table 4-15 list the top four projects and the worst four projects in the secondary evaluation. The project that received the highest total score is Research of Performance Assessment and Product Certification for Residential Building in China. The construction of residential buildings is in progress in China, which requires quality improvement with the establishment of a performance assessment system for housing and a certification system for the products and promoting a healthy residential building industry in order to develop an ethical and orderly housing market. The project aimed to establish a performance assessment system for housing and a certification system for housing products. A trial performance assessment system was implemented and institutional preparation is almost complete to implement a product certification system; both systems are scheduled to be fully implemented in 2005.

This project received 3.9 to 4.0 points for every evaluation criterion. The reasons for the high rating can be summarized as follows. First, “effectiveness” was rated high, because the establishment of methods for data collection and data examination regarding the safety and comfort level of housing that was achieved as an outcome contributed to the establishment of both of the aforementioned systems. Then, “efficiency” was marked high, because the specialization of dispatched experts and the timing of dispatch were appropriate and the equipment was locally procured. The achievement of the project purpose facilitated the achievement of the overall goal on the one hand, and the concept of “housing projects” consisting of multiple products was introduced and the technical standard of “housing products” was promoted by the project on the other hand, thus generating high “impacts.” Likewise, “sustainability” was evaluated high, because the certificate standard established by the project is under consideration to become a national accreditation, and it has been decided that the Chinese counterpart agency of the project will

Table 4-14 Ranking of the Top 4 Projects Rated “Excellent” by Secondary Evaluators

Country	Project Name	Total Score	Fiscal Year of Evaluation
1 China	Research of Performance Assessment and Product Certification for Residential Building	3.95	2004
2 Thailand	Project for Development of Trauma Center Complex	3.78	2004
3 Turkey	Project on Energy Conservation	3.74	2005
4 Thailand	Development of the Method of Urban Development	3.73	2004

Table 4-15 Ranking of the Worst 4 Projects Rated “Poor” by Secondary Evaluators

Country	Project Name	Total Score	Fiscal Year of Evaluation
1 Brazil	Technology Development for Revegetation and Utilization of Degraded Areas in the Semi-arid Region of the Northeastern Brazil	2.33	2004
2 Thailand	A Pilot Project to Construct a Recycling System in Southern Thailand	2.34	2005
3 Viet Nam	Strengthening of National Institute of Veterinary Research	2.48	2004
4 Cambodia	The Capacity Building for the Forestry Sector	2.48	2004

continue to play a leading role in the establishment of both systems.

In contrast, the project that received the lowest total score is Technology Development for Revegetation and Utilization of Degraded Areas in the Semi-arid Region of the Northeastern Brazil. In the semi-arid region of northeastern Brazil, where degraded land is rapidly expanding, this project aimed to develop afforestation technology and promote sustainable cattle production technology utilizing the afforestation technology for the purpose of recovering the degraded areas and preventing desertification. In this project, due to the late issuance of Brazilian visas to the experts, experts were dispatched more than one year after the start of the project. Because of this delay, the project purpose was not achieved and the forecast of the achievement of the overall goal was also difficult. Thus, the extension of cooperation period was recommended.

The highest score that the project received was 2.9 points for “relevance” criterion. The scores for “effectiveness,” “efficiency,” and “sustainability” were all around 2.3 points, and the one for “impact” did not even reach 2.0. The reasons for the overall low rating can be considered as follows. In terms of “effectiveness,” the delay in the dispatch of experts caused another delay in the actual start of the project by more than one year, and as a result, the project purpose was not achieved. In terms of “efficiency,” factors such as the delay in the dispatch of long-term experts caused by the external condition of delay in visa issuance, the delay in material and equipment procurement, and insufficient allocation of counterparts decreased efficiency. As for “impact,” because the project purpose was only partially achieved, the project was not at the stage of achieving the overall goal. Because the methods for promotion could not be identified, the project could not obtain the emergence of impacts.

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

“Relevance” of the target projects was generally high and the other factors, “effectiveness,” “impact,” and “sustainability,” achieved a certain level.

When “five evaluation criteria” were examined in terms of viewpoints, the “validity” of project implementation was high in terms of consistency with Japan’s aid policies, JICA Country Programs, the development policies of the partner countries, adequacy of the implementation as ODA, and priority of project implementation as cooperation in the partner country and sector. The “necessity” of the project was also high with regards to whether the project is in line with the needs of the target group, area, and society, and whether those needs are still present. On the other hand, the viewpoint of “appropriate approach” was marked relatively low with regards to whether the approach was appropriate as an effective solution to the development issues, whether the approach was appropriate to solve the preset development issue (overall goal), whether the selection of target area and group was appropriate, whether Japanese technology was superior, and

whether partnership with other donors and related projects in the partner country was planned and implemented to achieve higher level of outcomes.

As for “efficiency,” the viewpoint of “clear input cost” was rated relatively high: the viewpoint asks whether unit costs for purchasing equipment and dispatching experts were clearly presented. On the other hand, “cost-effectiveness” was rated low, less than 3.0 points: this viewpoint questions whether efforts for cutting costs were made (e.g. using local resources), whether there was any alternative means to reduce the cost, whether it was impossible to produce greater achievements at the same costs, and whether the cost-effectiveness was high compared to similar projects of other donors and/or the partner country.

As for “sustainability,” what were marked high scores were for the viewpoint of “organizational sustainability,” regarding the positioning of activities in the policies and whether organization of the implementing agency was stable enough for continuous effects, and the viewpoint of “technological sustainability,” asking whether technology and capacity acquired in the project were maintained and expanded and whether equipment was properly maintained and managed. What was rated low, on the other hand, was the viewpoint of “financial sustainability” regarding whether budget, e.g. operating expenses, was secured, and if the measures for budget were sufficient, with low scores falling below 3.0.

In spite of the limitation described above, when evaluation results were chronologically compared from 2003 to 2005, there were significant differences between 2003 and 2004 and between 2003 and 2005, showing that projects in fiscal 2004 and 2005 are higher in quality than projects in fiscal 2003. Analysis made last fiscal year also showed a significant difference between projects in fiscal 2003 and projects in fiscal 2004. Many projects in fiscal 2004 were evaluated this fiscal year so that it can be concluded that this year’s results are more stable and confirmed the results of the last fiscal year. However, no significant differences were observed between fiscal 2004 and fiscal 2005.

When the projects rated generally high and projects rated generally low by secondary evaluators were compared, there was a significant difference in each of the five evaluation criteria. Since the difference is especially large in “effectiveness,” “impact,” and “sustainability,” it is necessary to pay particular attention to these criteria when managing projects.

A comparative analysis was conducted by sector and region. The sectors of social development, health and medical care, and mining and industrial development received relatively higher scores for each criterion. Those three sectors tended to receive relatively lower scores for “effectiveness,” “efficiency,” and “impact” than those for “relevance” and “sustainability” on average. As for the agricultural development and forest and natural environment sectors, the average scores for most evaluation criteria are in the 2-point range and both sectors tended to receive relatively lower scores for “impact” and “sustainability” than those for “relevance,” “effectiveness,” and “efficiency” on aver-

age. With regard to regional comparisons, the Eastern Europe and the Middle East region were rated relatively high while Latin America was rated low. There was a similar tendency in the results for the Middle East and Eastern Europe region, and the Asia and Oceanic region. Every evaluation criterion was rated average or higher and “relevance” and “sustainability” were rated relatively high while “efficiency” was rated the lowest. The tendencies for Latin America and Africa were similar, with relatively high scores for “relevance” and “efficiency,” while the lowest scores were for “impact.” In the comparison by sector and region, only a small number of projects belong to a group sometimes, when projects with extreme outcomes more likely to affect the rating for the whole group. In order to minimize such noise and to obtain stable analysis results, more projects should be subject to the analysis.

When comparing the projects managed by JICA headquarters and those managed by overseas offices, a difference in the evaluation quality of terminal evaluation reports was confirmed, whereas there were no differences in the project evaluation by secondary evaluators. As for projects with ex-ante evaluation and projects without ex-ante evaluation, no differences were observed in the project evaluation by secondary evaluators either. Since the number of projects managed by overseas offices and projects with ex-ante evaluation subject to this year’s analysis was quite small, we have to wait for further detailed analysis with more such projects as samples.

Various factors are believed to influence the project outcomes. Among them, “project scale,” “cooperation period,” “GDP/cap,” “HDI,” and “education index” were selected to check the correlations among the criteria and viewpoints of the five evaluation. Some viewpoints in the criterion of “sustainability” had a strong correlation with GDP/cap, HDI and education index. Although these factors are external factors, not directly related to projects, it is clear that factors such as financial strength, organizational management capacity, governance, and general intelligence standard affect sustainability in a broad sense. It can be assumed that sustainability strongly depends on the efforts and mechanism of the partner country after the project is completed.

(11) For the Betterment of the Quality of Projects

In sum, some points for improving the quality of projects can be summarized as follows.

1) Appropriate Approach for Project Implementation

A project commences with an examination of the relevance, which is closely related to the subsequent implementation process and outcomes. The project evaluation by secondary evaluators for “relevance” resulted in relatively low scores for the viewpoint of “appropriate approach” compared to those for “validity” and “necessity.” However, as discussed in 1-3 (3), there is a strong correlation between “appropriate approach” and project outcomes and sustainability.

It is therefore necessary to select an appropriate approach

and method in implementing projects. Appropriate approach and method ensure the achievement of a project purpose, the generation of impacts, and an increase in sustainability, consequently leading to the improvement of quality of the whole project. Examining the appropriateness of approach at ex-ante evaluation is a key to the future success of the project.

2) Clear Causal Relationships 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 weak relationships between the project purpose and overall goal. In order to facilitate the outcome of the overall goal, it is necessary to further clarify the causal relationships among activities, project purpose, and overall goal. It is also important to set an appropriate overall goal in consideration of the fact that the project may lose its direction if the overall goal deviates too much from the project purpose.

3) Setting Indicators and Numerical Targets

Some projects set numerical targets that are unlikely to be achieved; for example, indicators to measure the level of achievement of the project activities and purpose were not set, or no clear numerical targets (parameter) 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 numerical target in order to conduct an objective verification.

4) Utilization of PDM

Some projects that utilized the PDM for monitoring purposes were generally rated high at the secondary evaluation. Utilization of PDM enables the concerned parties to be aware of what needs to be done and how in order to achieve the purposes/goals, thus contributing to the facilitation of project outcomes.

5) Better Understanding of PDM among Concerned Parties

It turned out that in some projects there was a discrepancy in understanding with regard to the project purpose between Japan and the counterparts. It is necessary to avoid abstract expressions and use clear wordings in the definition of words used for PDM. It is also necessary for both parties to fully understand the project purpose and process leading to the achievement of the purpose by communicating with each other sufficiently.

6) Response to Changes of External Conditions

In some projects, external conditions such as a change in the political climate of the partner country significantly affected the progress of the project. When a significant change in external conditions is observed, it is necessary to take measures such as significantly modifying the project contents based on the mid-term evaluation instead of continuing the project as it is.

Appendix 1 List of Projects Subject to Secondary Evaluation in Fiscal 2006

Fiscal 2004 (New Targets): 28 projects	
Argentina	Project on Establishment of Control Capacity for Industrial Wastewater and Waste
Bangladesh	Mobile Arsenic Center for Mitigation of the Arsenic Contamination of Drinking Water
Brazil	Technology Development for Revegetation and Utilization of Degraded Areas in the Semi-arid Region of the Northeastern Brazil
Cambodia	The Capacity Building for the Forestry Sector
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	Human Resource Development Project for Water Resources
China	Research of Performance Assessment and Product Certification for Residential Building
Egypt	The Water Management Improvement Project in the Nile Delta
Fiji	Information and Communication Technologies (ICTs) Capacity Building at the University of the South Pacific
Ghana	Improvement of Educational Achievement in Science, Technology and Mathematics (STM) in Basic Education
Ghana	The Small-scale Irrigated Agriculture Promotion Project (Follow-up)
Hungary	Human Resources Development for Environmental Engineers at the College of Dunaujvaros
Indonesia	Human Resources Development for Local Governance
Indonesia	Strengthening of Polytechnic Education in Electric-related Technology
Iran	Project on the Improvement of Audio-visual Aids and Instruction Methods in Vocational Training at the Instructor Training Center (ICT)
Mexico	Development of Method of Research and Education in Electric Field
Paraguay	Improvement of Small and Medium Scale Dairy Farm Management Project
Philippines	Project for Strengthening of Continuing School Based Training Program for Elementary and Secondary Science and Mathematics Teachers (SBTP-ELSSMAT)
Sri Lanka	Project for Human Resource Development in Information technology through Capacity Building of University of Colombo School of Computing
Thailand	Regional Cooperation Project on Capacity Building of Drug Analysis for Improvement of Drug Law Enforcement in Thailand, Cambodia, Lao P.D.R., Myanmar and Viet Nam
Thailand	Development of the Method of Urban Development
Thailand	Project for Development of Trauma Center Complex
Thailand	The Project on the Practical Energy Management Training Center
Thailand	Project on Local Management Cooperation
Tunisia	The Project for Strengthening of Reproductive Health Education
Viet Nam	Strengthening of National Institute of Veterinary Research
Viet Nam	The Project for Improvement of Cattle Artificial Insemination Technology
Fiscal 2005 (New Targets): 17 projects	
Bangladesh	The Project to Set Up Rural Development Engineering Center
China	Water Environment Restoration Pilot Project in Taihu Lake
Indonesia	The Forest Fire Prevention Management Project (Phase 2)
Indonesia	The Demonstration Study on Carbon Fixing Forest Management
Indonesia	Establishment and Capacity Building of Regional Export Training and Promotion Centers
Kazakhstan	Technical Cooperation for the Improvement of Health Care Services in the Semipalatinsk Region
Laos	Development of Bases to Autonomously Carry out Reading Promotion Project
Malaysia	Project on Networked Multimedia Education System
Myanmar	Project for Primary Health of Mothers and Children
Paraguay	Strengthening Continuing Education in Nursing and Midwifery in the South of the Republic of Paraguay
Philippines	Water Buffaloes and Beef Cattle Improvement Project
Thailand	A Pilot Project to Construct a Recycling System in Southern Thailand
Thailand	The Project on Capacity Building for Environmental Management and Airside Paved-area Maintenance of Suvarnabhumi Airport
Thailand	Development Vocational Opportunities and Creative Activities for People with Disabilities and Commercializing Hill-tribes Peoples' Crafts
Thailand	Project on Developing the Capacity of the Government to Post Evaluate the Externally Funded Project
Tunisia	Project for the Establishment of the Vocational Training Center for the Electric and Electronics Industry
Turkey	Project on Energy Conservation
Both Years as Seam Allowance (2004 and 2005): 15 projects	
Brazil	The Technological Development Project for Sustainable Agriculture in Eastern Amazonia
Chile	Improvement of Productivity for the Small-scale Dairy Farmers Project
China	Anhui Primary Health Care Technical Training Center
China	Research Project on Timber from Man-made Forests
Kenya	Promotion of Sustainable Community Based Small-holder Irrigation
Malaysia	The Project for the Development of Technology Related to the Processing of Feed Based on Agro-industrial By-products of Oil Palms Production (Follow-up)
Malaysia	The Project for Strengthening of Food Safety Programme
Nepal	Community Development and Forest/Watershed Conservation Project (Phase 2)
Senegal	High-level Technician (BTS) Training Project at the Senegal-Japan Vocational Training Center
Philippines	Environmental and Productivity Management of Marginal Soils
Philippines	Promotion of the Ship Inspection System and Technique
Thailand	The Reforestation and Extension Project in the Northeast of Thailand (Phase 2)
Viet Nam	Project on the Improvement of Higher Maritime Education
Viet Nam	The Project for Strengthening Training Capability for Technical Workers in the Hanoi Industrial College
Zambia	Technical and Vocational Training Improvement Project (Aftercare)

Appendix 2

Secondary Evaluation Check Sheet (Fiscal 2006)

Evaluator		Date
Project title		

Rating criteria

<p>1) Rate viewpoints and criteria in green cells and orange cells based on a scale of 1 to 5. [I – III]</p> <p>5: Sufficient/high 4: Fairly sufficient/high 3: Average 2: Slightly insufficient/low 1: Insufficient/low *: Cannot tell</p> <p>2) Rate familiarity in green cells choosing from the dropdown list.</p> <p>3) Write down highlights and notable points (including good practices) in the space for comment. [I – IV]</p>
--

I. Preconditions for appropriate evaluation (evaluability)

[Appropriateness of Project Plan (Preliminary Study or PDM)] Whether project plan (preliminary study/PDM) is designed properly?		Rating
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]		Rating
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?		Rating
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?		Rating
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?		Rating
Viewpoint	The PDM for the evaluation sets a clear and realistic hypothesis and logical flow considering important external assumptions.	
[Verifiability of Performance and Implementation Process] Whether project monitoring is conducted and documentation is properly conducted?		Rating
Viewpoint	Monitoring of outputs, activities, and inputs was regularly conducted, and the information including statistical data is accumulated during project implementation.	
Rating	Comment	Overall

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.		Rating
Viewpoint	Whether it is clearly mentioned in the report about the concerns for the neutrality/fairness/expertise of manpower resource. Or, (if not,) based on the 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).		Rating
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	Overall

2. Data Collection*¹		
[Evaluation Questions] Whether evaluation questions are set properly.		Rating
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.		Rating
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.		Rating
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.		Rating
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	Overall
3. Analysis		
3.1 Assessment of Performance and Verification		
[Comprehension and Verification of Project Performance (Outputs)] Whether outputs are comprehended and verified sufficiently.		Rating
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.		Rating
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.		Rating
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.		Rating
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)]		Rating
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)]		Rating
Viewpoint	The causal relationships are thoroughly examined to verify that effects for the beneficiaries have resulted from the project interventions* ³ .	
Rating	Comment	Overall
3.2 Analysis Method		
[Objective Analysis] Whether objective analysis is conducted based on data.		Rating
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.		Rating
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/impeding factors is conducted.		Rating
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	Overall

4. Evaluation by Five Criteria*4		
[Five Evaluation Criteria (Relevance)] Whether the evaluation on relevance is sufficient.		Rating
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.		Rating
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.		Rating
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.		Rating
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.		Rating
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.		Rating
Viewpoint	The conclusion is drawn based on holistic viewpoints that are in turn based on the five evaluation criteria.	
Rating	Comment	Overall
5. Recommendations/Lessons Learned*5		
5.1 Recommendations		
[Recommendations (Sufficiency of Recommendations)] Whether recommendations are drawn sufficiently.		Rating
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.		Rating
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.		Rating
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	Overall
5.2 Lessons Learned		
[Lessons (Sufficiency of Lessons Learned)] Whether lessons are fully drawn.		Rating
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 drawn from the evaluation result and include persuasive contents.		Rating
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.		Rating
Viewpoint	The lessons are generalized and conceptualized so that they are widely applicable in the future.	
Rating	Comment	Overall

6. Reporting		
[Presentation (Conciseness, Clarity, Clearness)] Whether the report is presented in a concise and clear manner so that the readers comprehend easily.		Rating
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.		Rating
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.		Rating
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	Overall

III. Evaluation of the Project Based on the Report

Fill in comments if there are any external 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.		Rating
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.		Rating
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.		Rating
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	Overall
2. Effectiveness (Achievement of Project Purpose)		
[Achievement Level of Project Purpose] Whether project purpose is achieved.		Rating
Viewpoint	Project purpose has been (is going to be) achieved.	
[Causal Relationships between Outputs and Project Purpose] Whether cause-effect relationship is strong enough.		Rating
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 inhibited or promoted effectiveness.	
Rating	Comment	Overall
3. Efficiency (Efficiency of Project)		
[Clarity of Input Cost] Whether input cost is comprehended clearly.		Rating
Viewpoint	Unit costs for purchasing equipment and dispatching experts are clearly presented.	
[Cost-effectiveness] Whether utmost efforts are made for cost-effectiveness.		Rating
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.		Rating
Viewpoint	The inputs were made in a timely manner with appropriate scale and quality.	
Rating	Comment	Overall

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.		Rating
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.		Rating
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.		Rating
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	Overall
5. Sustainability (Sustainability after Completion of JICA's Technical Cooperation)		
[Mechanism of Securing Sustainability] Whether mechanism for sustainability are institutionalized through project implementation.		Rating
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.		Rating
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.		Rating
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.		Rating
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.		Rating
Viewpoint	Budget including operating expenses is secured. Measures for securing budget are sufficient.	
Rating	Comment	Overall

IV. Familiarity toward the Concerned Project

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