

**Third party Quality Check for  
Ex-post Evaluation  
(Internal Evaluation) in JICA  
  
Analysis Report in 2020**

Consultant: KPMG AZSA LLC

# Targeted projects of third party QC in 2020

- Number of targeted projects: 52 (integrated evaluation (TC): 10, Integrated evaluation (TC & GA): 2)  
Regarding Integrated evaluation (TC & GA), it is counted as 2 because 2 third party QCs were performed for TC and GA. Therefore, the number of evaluation votes is 51.
- Technical Cooperation: 42 (including 10 integrated evaluations (TC), 1 integrated evaluation (TC & GA), and 6 SATREPS (note)), Technical Cooperation for Development Planning: 5, Grant Aid: 5 (including 1 integrated evaluation (TC & GA))
- Overall Evaluation: Highly satisfactory: 7, Satisfactory: 30 (integrated evaluation (TC): 7, integrated evaluation (TC & GA): 2), partially satisfactory: 13 (integrated evaluation (TC): 2), Unsatisfactory: 2 (integrated evaluation (TC): 1)

The population (projects for which an internal ex-post evaluation was conducted in FY2019) is as small as 100 or less. Therefore, when extracting the target projects, we extracted significant projects that require confirmation of the quality of evaluation, rather than random sampling. We paid attention to the balance of sectors and schemes. (This report describes the secular changes of the FY2018, FY2019, and FY2020. However, the selection criteria for the target projects are different, therefore, it is necessary to note that simple comparison is not always possible.)

## List of the targeted project

No.	Country	Name of the project	Scheme	Note
1	Indonesia	Project on Capacity Development of Animal Health Laboratory	TC	
2	Cambodia	Project on Improving Official Statistics in Cambodia (Phase 2) (Phase 3)	TC	Integrated evaluation
3	Cambodia, Laos, and Vietnam	Project for the Capacity Development for Transition to the New CNS/ATM Systems in Cambodia, Lao PDR, and Viet Nam	TC	
4	Vietnam	Capacity development for NIHE to control emerging and re-emerging infectious diseases (Phase 1) Project for Capacity Development for Laboratory Network in Vietnam of Biosafety and Examination of Highly Hazardous Infectious Pathogens (Phase 2)	TC	Integrated evaluation
5	Vietnam	Sustainable Integration of Local Agriculture and Biomass Industries	TC	SATREPS
6	Myanmar	Project for Improvement of Medical Equipment in General Hospitals in Yangon	GA	

# List of targeted project

No.	Country	Name of the project	Scheme	Note
7	Philippines	Project for Prevention and Control of Leptospirosis in the Philippines	TC	SATREPS
8	Papua New Guinea	Project for Promotion of Smallholder Rice Production (Phase 1) (Phase 2)	TC	Integrated evaluation
9	Timor-Leste	Irrigation and Rice Cultivation Project in Manatuto (Phase 1) (Phase 2)	TC	Integrated evaluation
10	China	Environment Construction at Co-existent Areas of Human Beings and Crested Ibis	TC	
11	China	Project on Capacity Building for Occupational Health	TC	
12	Armenia	Project for Development of Local Production and Promotion of Local Brands	TC	
13	India	The Study on Development and Management of Land and Water Resources for Sustainable Agriculture in Mizoram	TCDP	
14	Nepal	Strengthening the Monitoring and Evaluation System in Nepal Project for Strengthening the Monitoring and Evaluation System in Nepal Phase 2	TC	integrated evaluation
15	Afghanistan	National Agricultural Experiment Stations Rehabilitation Project (NARP)	TC	
16	Afghanistan	Rice-based agriculture in Nangarhar Province (RIP Project)	TC	
17	Sri Lanka	Urban Transport System Development Project for Colombo Metropolitan Region and Suburbs	TCDP	
18	Brazil	Social Inclusion through the Incentive to Produce Oleaginous Plants for the Generation of Bio-diesel in the State of Rio Grande do Norte	TC	
19	Brazil	Regional Assessment of Carbon Dynamics in the Amazonian Forests	TC	SATREPS
20	Peru	Project for Improvement of Equipment for Disaster Risk Management	GA	
21	Bolivia	Project for Drink Water Provision in Rural Area of Beni & Pando Prefectures in Bolivia	GA	
22	Ethiopia	The One Village One Product Promotion Project	TC	

# List of targeted project

No.	Country	Name of the project	Scheme	Note
23	Zambia	Health Capital Investment Support Project	TC	
24	Malawi	Project for Enhancement of Operation and Maintenance for Rural Water Supply	TC	
25-1	Democratic Republic of the Congo	Project on Capacity Development for Bridge Management	TC	Integrated evaluation (TC & GA)
25-2	Democratic Republic of the Congo	Project of Improvement of the Marshal Bridge in Matadi	GA	Integrated evaluation (TC & GA)
26	Senegal	Project for Sanitation and Hygiene Improvement in Rural Areas of Tambacounda, Kedougou, and Matam Regions	TC	
27	Egypt	Project for Sustainable Systems for Food and Bi-energy Production with Water-saving irrigation in the Egyptian Nile Basin	TC	SATREPS
28	Iraq	Project on Master Plan Study for Port Sector in Iraq	TCDP	
29	Iran	Establishment of Emergency Response Plan for the first 72 Hours after Earthquake (Phase 1) Capacity Building for Earthquake Risk Reduction and Disaster Management in Tehran (Phase 1)	TC	Integrated evaluation
30	Turkey	Industrial Automation Technology (IAT) Extension Project for Central Asian/Middle East Countries	TC	
31	Bosnia-Herzegovina	Project for Herzegovina International Tourism Corridor Development and Environmental Conservation	TC	
32	Palestine	Strengthening Support System focusing on Sustainable Agriculture in Jericho and Jordan River Rift Valley Project on Improved Extension for Value-Added Agriculture in the Jordan River Rift Valley	TC	Integrated evaluation

# List of targeted project

No.	Country	Name of the project	Scheme	Note
33	Guatemala	Project for the Capacity Development of Local Governments	TC	
34	Indonesia	Project on Building Administration and Enforcement Capacity Development For Seismic Resilience (Phase 1 & 2)	TC	Integrated evaluation
35	Indonesia	Project on Mangrove Ecosystem Conservation and Sustainable Use in the ASEAN Region	TC	
36	Thailand	Project for Research and Development for Waste Reuse Technology in Tropical Region	TC	SATREPS
37	China	Project for Capacity Building of Reproductive Health and Family Care Service in Central and Western Region	TC	
38	China	Project for Strengthening of Health Education for Prevention of Infectious Diseases through Family Health	TC	
39	Nicaragua	Strengthening of Activities of Survey and Control for Chagas Disease	TC	
40	Nicaragua	Project for the Study of National Transport Plan in the Republic of Nicaragua	TCDP	
41	Kenya	Project for Sustainable Smallholder Irrigation Development and Management in Central and Southern Kenya (SIDEMAN)	TC	
42	Ghana	Project for Strengthening Operational Capacity of Prevention of Mother-to-Child Transmission of HIV (PMTCT)	TC	
43	Sierra Leone	Sustainable Rice Development Project in Sierra Leone	TC	
44	Tanzania	Project for Institutional Capacity Strengthening for HIV Prevention Health System Strengthening for HIV and AIDS Service Project	TC	Integrated evaluation
45	Mozambique	Chokwe District Irrigation for Small Farmer's Comprehensive Agricultural Development Project Project for Rice Productivity Improvement in Chokwe Irrigation Scheme	TC	Integrated evaluation
46	Mozambique	Project for Enhancement of the Capacity of Destination Marketing and Promotion through Strengthening the Linkage among Tourism Related Organizations	TC	

# List of targeted project

No.	Country	Name of the project	Scheme	Note
47	Burkina Faso	Digital Topographic Mapping Project in Burkina Faso	TCDP	
48	Tunisia	Project for the Development of Irrigated Areas of Northern Tunisia	TC	
49	Croatia	Project on risk identification and land-use planning for disaster mitigation of landslide and floods	TC	SATREPS
50	Uganda	Project for Provision of Improved Water Source for Resettled Internally Displaced Persons in Acholi Sub-region	GA	
51	Fiji	Waste Minimization and Recycling Promotion Project	TC	

note: SATREPS means "Science and Technology Research Partnership for Sustainable Development".

# Objective and background

JICA evaluates projects for mainly fulfilling two purposes: 1) improve the projects further through the PDCA cycle and 2) ensure accountability to stakeholders including Japanese nationals and people of the partner. Internal ex-post evaluation is conducted by JICA overseas office staff and regional departments in the headquarter. The targets of internal ex-post evaluation are the projects which cost 200 million yen or more and under one billion yen.

In order to achieve the objectives of evaluation more effectively and efficiently, the internal evaluation process is being improved by introducing self-assessment by internal evaluators and quality check by external third parties.

The objectives of this study is to examine internal evaluations by external third party and to ensure the quality, objectivity, and impartiality of future internal evaluations as well as to disclose its outline to the public to strengthen the accountability.

## Third party QC tools and information collection methods

1. Third party QC is conducted through the desktop research. External third parties examine internal evaluation report and verify them based on a check sheet similar to the one developed for self-assessment.
2. The person in charge of third party QC ticks Yes/ Partly Yes/ No/ Not Applicable for each checklist item and comment the reasons on the check sheet.
3. Overall trends (e.g. average and characteristic results etc.) are analyzed based on the result of third party QC of individual projects.
4. For analysis of the overall trends, the results of the third party QC are calculated as score.
5. Information source for third party QC is limited to the internal evaluation report in principle.
6. Third party QC check sheet is aligned with the JICA Guidelines for Project Evaluation and Internal Evaluation Implementation Manual.
7. This year is the 3rd third party QC, following 2018 and 2019. After the completion of the third party QC in each year, the checklist is partially revised and the scoring criteria are clarified based on the experience. This report describes changes over time since 2018, but the scoring criteria have been slightly revised. Therefore, it should be noted that simple comparison is not always possible.

# Methodology of scoring

## ① **Raw score :**

Score 2 points for “Yes”, 1 point for “Partly Yes”, 0 point for “No”, and no score for “Not Applicable”.

## ② **Standardized score :**

Standardized score = total of raw score ÷ (total number of checklist items – number of tick on “Not Applicable” )×2

Standardized score is between 0.0 to 1.0. As the score is closer 1, it means that there are more ticks on “Yes”, in other words, the internal evaluation’s quality is higher.

[Note : Calculation of dispersion and standard deviation]

$$\text{Dispersion : } \frac{1}{n} \sum_{n=1}^n (x_i - \bar{x})^2 \quad \text{Standard deviation : } \sigma = \sqrt{\frac{1}{n} \sum_{n=1}^n (x_i - \bar{x})^2}$$

n = total number of targeted projects、 $x_i$  = average of standardized score、 $\bar{x}$  = standardized score

The dispersion and standard deviation were calculated for the whole and for each evaluation criteria, and the variation in the third party QC results of the target project was confirmed. The larger the dispersion / standard deviation, the greater the variation from the mean score.

# Items to be analysed

1. Overall trends
2. Evaluation criteria (①Relevance, ②Effectiveness / Impact, ③Efficiency, ④Sustainability, ⑤Conclusion, Recommendations and Lessons Learned, ⑥General Matters)





Third party QC of target projects



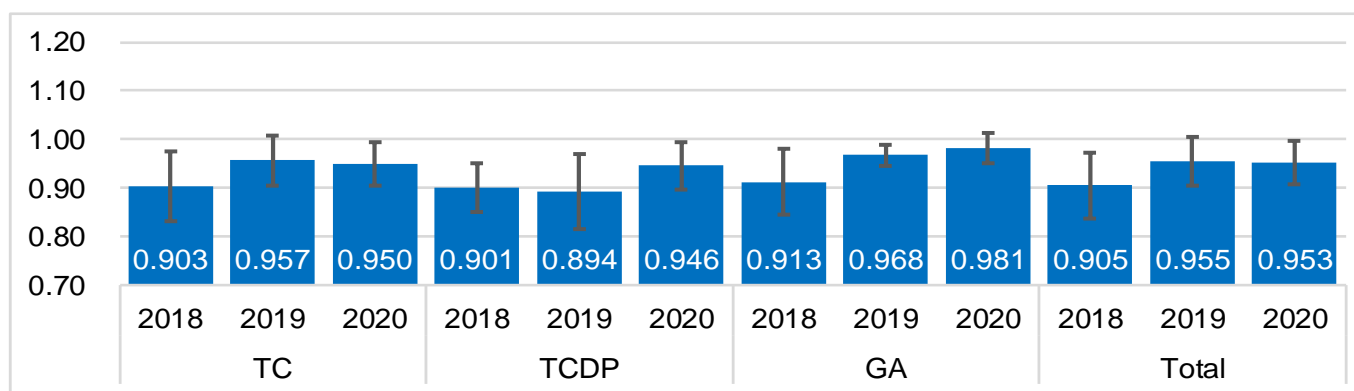
# 1. Overall trend

# 1. Overall trend

Average, dispersion, standard deviation by standardized score of 52 target projects (and 2018 and 2019 results)

	TC			TCDP			GA			Total		
Year	2018	2019	2020	2018	2019	2020	2018	2019	2020	2018	2019	2020
Average	0.903	0.957	0.950	0.901	0.894	0.946	0.913	0.968	0.981	0.905	0.955	0.953
Dispersion	0.005	0.003	0.002	0.003	0.006	0.002	0.004	0.000	0.001	0.005	0.003	0.002
Standard deviation	0.072	0.051	0.045	0.051	0.077	0.048	0.067	0.022	0.031	0.068	0.051	0.044

Note: In FY2018, TCDP was included in TC in the data aggregation. Since TCDP is calculated separately from FY2019, the data for FY2018 is also calculated again.



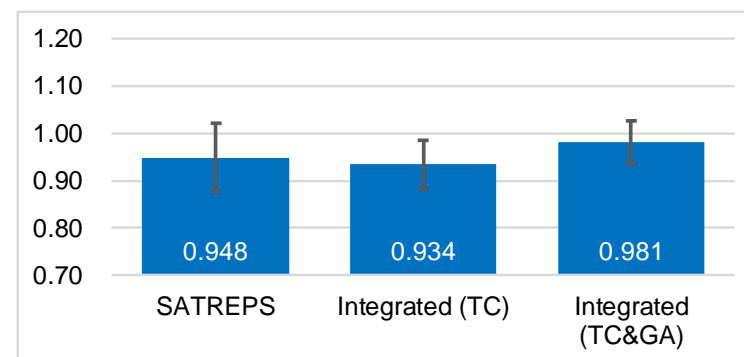
Note: Bar graph: average of standardized score, error bar (error range):  $\pm$  standard deviation

- As an overall level, the average third party QC standardized score in 2020 was slightly above 0.950. It was almost the same level as in 2019.
- By scheme, GA had a slightly higher average score than other schemes.

# 1. Overall trend

Average, dispersion, standard deviation by standardized score of target projects (6 SATREPS, 10 integrated evaluation (TC) and 2 integrated evaluation (TC & GA)) (and 2018 and 2019 results)

	SATREPS	Integrated evaluation (TC)	Integrated evaluation (TC & GA)
Average	0.948	0.934	0.981
Dispersion	0.003	0.003	0.001
Standard deviation	0.052	0.055	0.026



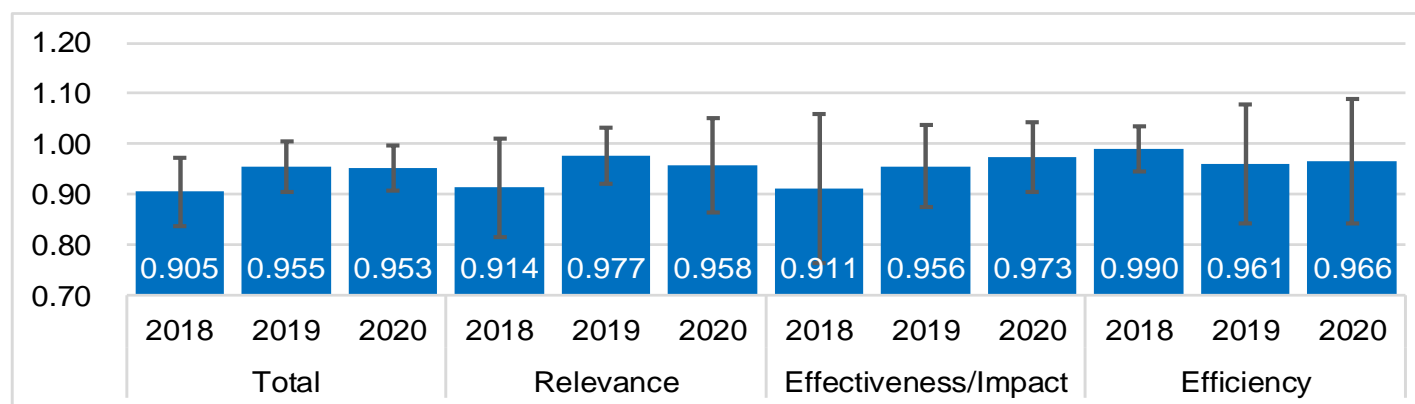
Note: Bar graph: average of standardized score, error bar (error range):  $\pm$  standard deviation

- The average standardized score for integrated evaluation (TC & GA) is high, but the number of evaluation was only one, so it can be said that the score depends on the characteristics of individual projects.
- The average standardized score of integrated evaluation (TC) is a little low. When the average score for each evaluation criteria is verified, average score of "relevance" (0.915), "effectiveness / impact" (0.916), "conclusion / recommendation / lessons learned" (0.883) are low.
- The average standardized score for SATREPS was about the same as the overall average score.

# 1. Overall trend

Average, dispersion, standard deviation by standardized score of 52 target projects by evaluation criteria (and 2018 and 2019 results)

	Total			Relevance			Effectiveness / Impact			Efficiency		
	2018	2019	2020	2018	2019	2020	2018	2019	2020	2018	2019	2020
Average	0.905	0.955	0.953	0.914	0.977	0.958	0.911	0.956	0.973	0.990	0.961	0.966
Dispersion	0.005	0.003	0.002	0.009	0.003	0.009	0.022	0.006	0.005	0.002	0.014	0.015
Standard deviation	0.068	0.051	0.044	0.097	0.056	0.093	0.148	0.081	0.069	0.044	0.117	0.122

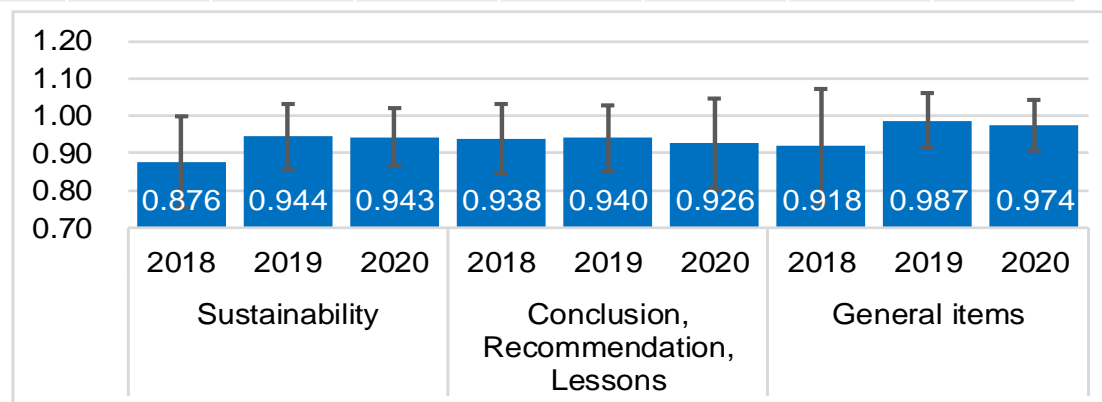


Note: Bar graph: average of standardized score, error bar (error range):  $\pm$  standard deviation

# 1. Overall trend

Average, dispersion, standard deviation by standardized score of 52 target projects by evaluation criteria (and 2018 and 2019 results)

	Sustainability			Conclusion / recommendation / lesson learned			General items		
	2018	2019	2020	2018	2019	2020	2018	2019	2020
Average	0.876	0.944	0.943	0.938	0.940	0.926	0.918	0.987	0.974
Dispersion	0.015	0.008	0.006	0.008	0.008	0.015	0.024	0.005	0.005
Standard deviation	0.124	0.089	0.077	0.092	0.088	0.121	0.155	0.074	0.067



Note: Bar graph: average of standardized score, error bar (error range):  $\pm$  standard deviation

- The average standardized scores are high for "effectiveness / impact" (0.973) and "general items" (0.974). Regarding the evaluation of "effectiveness / impact", although there are differences for each check item, it seems that the quality of the evaluation analysis is generally ensured.
- The average standardized score is low for "conclusions / recommendation/ lesson learned" (0.926). Some of the evaluations are scored as insufficient or inappropriate in the content of the recommendations and lessons learned.
- "Efficiency" and "conclusions / recommendation / lesson learned" have slightly larger standard deviations (0.122 and 0.121). The reason for the large variation is that the number of check items is small and there are many evaluations scored as "not applicable".

# 1. Overall trend

### Correlation of standardized score of each evaluation criteria to the whole

For the 52 evaluations, the correlation coefficient of the sequence of standardized score of each evaluation criteria was calculated for the sequence of standardized score of the whole. The correlation coefficient indicates how much the quality and accuracy of the ex-post evaluation in each criteria correlates with the overall quality and accuracy.

	Correlation coefficient for standardized score (to the whole)	Number of third party QC check items
Total	1.000	TC: 33, GA: 33, TCDP: 31
Relevance	0.469	TC: 5, GA: 5, TCDP: 5
Effectiveness / Impact	0.404	TC: 8, GA: 7, TCDP: 6
Efficiency	0.185	TC: 3, GA: 4, TCDP: 3
Sustainability	0.775	TC: 10, GA: 10, TCDP: 10
Conclusion / recommendation / lesson learned	0.414	TC: 4, GA: 4, TCDP: 4
General items	0.261	TC: 3, GA: 3, TCDP: 3

[Level of correlation coefficient]

0.7-1 fairly strong correlation

0.4-0.7 strong correlation

0.2-0.4% weak correlation

0-0.2 almost no correlation

[Calculation formula of correlation coefficient (reference)]

Correlation coefficient between x and y =  

$$\frac{\text{Covariance of x and y}}{(\text{standard deviation of x}) \times (\text{standard deviation of y})}$$

- A positive correlation of "weak correlation" or higher (correlation coefficient of 0.2 or higher) is observed in all evaluation items except "efficiency". Due to the characteristics of the aggregation method, as a general tendency, the evaluation criteria with many check items have a larger correlation coefficient. It is probable that the reason why the correlation coefficients of "efficiency" and "general items" are low was that there are few check items.
- A "fairly strong correlation" is found for "sustainability". In addition to the many check items, the following factors are also assumed. "Sustainability" requires persuasive analysis centered on qualitative information, but on the other hand, analysis methods and judgment criteria are often left to the evaluator, making evaluation judgment difficult. Therefore, evaluators who perform high-quality (or low) analysis in sustainability perform high-quality / accurate (or low) analysis as a whole, and as a result, the correlation coefficient is higher than other evaluation criteria.

## 1. Overall trend

- The average of the 52 standardized scores is 0.953. It is about the same level as 0.955 points in 2019. Continuing from 2018 and 2019, the third party QC scoring results are high, and almost all check items are judged as "yes".
- By scheme, the average standardized score of GA is higher than others. However, this year, TC accounted for 42 of the 52 target projects, and TCDP and GA are 5 projects each, and most of the targets are TC. The high average score for GA seems to reflect the characteristics of individual projects rather than the characteristics of the scheme.
- The average standardized scores are high for "effectiveness / impact" (0.973) and "general items" (0.974). On the other hand, the average score of "conclusions / recommendation / lesson learned" is low (0.926). Trends and issues for each individual check item will be described later.

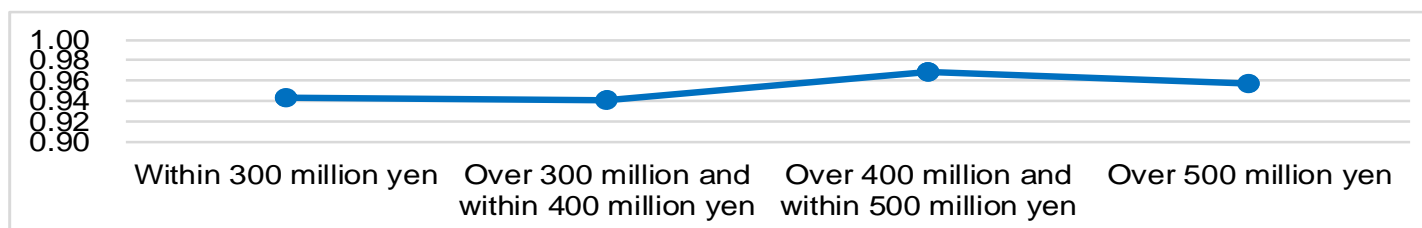


# 1. Overall trend

The 52 target projects are classified by project cost, elapsed period after project completion, and evaluation results, and the average standardized score are calculated for each category. In addition, each scheme is aggregated together.

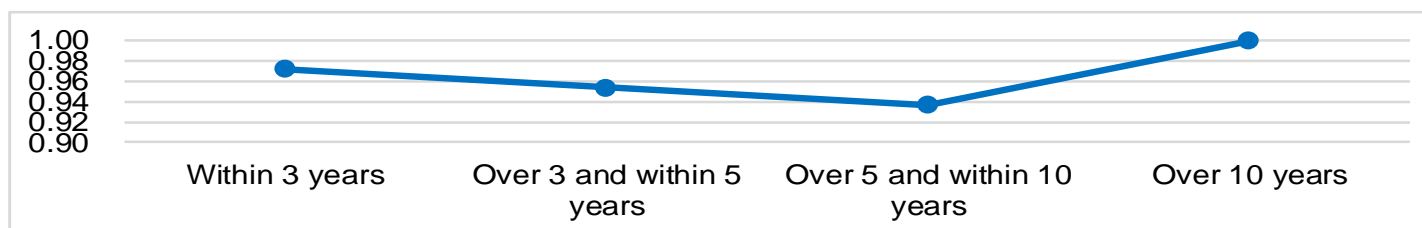
## Project cost

	Within 300 million yen	Over 300 million and within 400 million yen	Over 400 million and within 500 million yen	Over 500 million yen
Number of project	9	13	9	21
Standardized score (average)	0.943	0.941	0.969	0.957



## Elapsed period after project completion

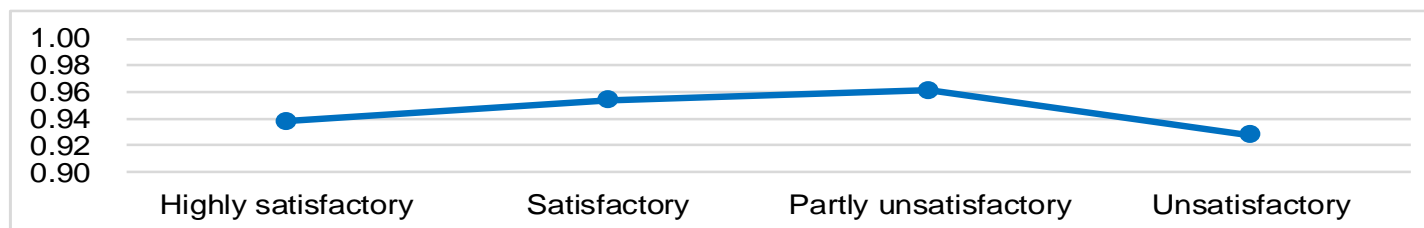
	Within 3 years	Over 3 and within 5 years	Over 5 and within 10 years	Over 10 years
Number of project	2	41	8	1
Standardized score (average)	0.972	0.954	0.937	1.000



# 1. Overall trend

## Evaluation result

	Highly satisfactory	Satisfactory	partially satisfactory	Unsatisfactory
Number of project	7	30	13	2
Standardized score (average)	0.938	0.954	0.961	0.928



- Regarding project costs, no tendency is observed for each category.
- Regarding the elapsed period after the completion of the project, there was only one case over 10 years, but excluding this, the average of the standardized score is higher when the elapsed period after the completion of the project is shorter. The shorter the period after the completion of the project, the easier it is to obtain information, and the quality and rationality of the evaluation may have improved.
- Regarding the classification of evaluation results, the average of the standardized score of “highly satisfactory” and “unsatisfactory” are slightly lower. However, as for the evaluation results, 30 out of 52 cases were concentrated on “satisfactory”, so the number of “highly satisfactory”, “partially satisfactory” and “unsatisfactory” are small, and the difference in average score was small. It seems that it is due to the influence of the results of individual projects rather than the characteristics based on the evaluation results.

# 1. Overall trend

How much difference there is in the standardized score of the self-assessment result and the third party QC result in each evaluation criteria of “relevance”, “effectiveness / impact”, “efficiency”, “sustainability”, “conclusion / recommendation / lesson learned”, and “general items” is confirmed and analyzed.

**Comparison of third party QC results and self-assessment results of standardized score of 52 target projects**

	Total	Relevance	Effectiveness / Impact	Efficiency	Sustainability	Conclusion / recommendation / lesson learned	General items
Third party QC	0.952	0.958	0.973	0.966	0.943	0.926	0.974
Self-assessment	0.899	0.944	0.933	0.841	0.877	0.955	0.864

- In the total, the standardized score of third party QC exceeds the standardized score of self-assessment . It can be said that the third party QC highly evaluates the quality of the evaluation.
- However, unlike self-assessment, it is difficult for third party QC to make a subjective judgment, and the focus is on format and appearance check, so it seems that the score tended to be slightly higher as a result. Due to the scoring rules, the criteria and way of thinking of scoring are slightly different, such as the cases where the third party QC marks "not applicable" but the self-assessment scores “partially yes” and “no”. This also seems to be the cause of the difference.
- In terms of “relevance,” “effectiveness / impact,” “efficiency,” “sustainability,” and “general items,” the standardized score for third party QC exceeds the standardized score for self-assessment.
- Only in “conclusion / recommendation / lesson learned”, the standardized score of third party QC was lower than the standardized score of self-assessment. In confirming the relationship between the contents in the evaluation report and the recommendations / lessons, and the usefulness of the recommendations / lessons, it seems that the third party QC gave a slightly stricter score than self-assessment. In the third party QC, it is pointed out that the contents described in the recommendations and lessons are not mentioned in the evaluation report. In some cases, the necessity, usefulness, and concreteness of recommendations and lessons are also pointed out.

# 1. Overall trend

### Difference analysis between third party QC and self-assessment

Third party QC check items	Number of items
Third party QC = Self-assessment	1,025 (60.1%)
Third party QC > Self-assessment	151 (8.9%)
Third party QC < Self-assessment	83 (4.9%)
The number of items where "not applicable" is checked by self-assessment and/or third party QC, the number of items in the third party QC that are not included in the self-assessment items, and the number of items not answered by Self-assessment	447 (26.2%)

Note: The population was 42 TC x 33 check items + 5 TCDP x 31 check items + 5 GA x 33 check items = 1,706 items. The ratio is rounded to the first decimal place.

- Approximately 60% of the check items are third party QC = self-assessment, and less than 30% are "not applicable" or check items that have no questions in the self-assessment or are not answered in the self-assessment.
- The number of check items with third-party QC > self-assessment is less than 10%, and the number of check items with third party QC < self-inspection was less than 5%, with the former having more cases.

# 1. Overall trend

Evaluation criteria	Difference between self-assessment and third party QC	Number of items (rate to the population)	Comments
Relevance (42 TC X 5 check items X + 5 TCDP X 5 check items + 5 GA X 5 check items = 260)	Third party QC = self-assessment	<b>186 (71.5%)</b>	● The results of self-assessment and third party QC are generally consistent.
	Third party QC > self-assessment	11 (4.2%)	
	Third party QC < self-assessment	12 (4.6%)	
	NA, with no items, with no answer	51 (19.6%)	
Effectiveness / impact (42 TC X 8 check items X + 5 TCDP X 6 check items + 5 GA X 7 check items = 401)	Third party QC = self-assessment	231 (57.6%)	● In many evaluations, third party QC scores higher than self-assessment.
	Third party QC > self-assessment	<b>28 (7.0%)</b>	
	Third party QC < self-assessment	<b>10 (2.5%)</b>	
	NA, with no items, with no answer	132 (32.9%)	
Efficiency (42 TC X 3 check items X + 5 TCDP X 3 check items + 5 GA X 4 check items = 161)	Third party QC = self-assessment	50 (31.1%)	● The number of evaluation checked as "Not applicable" is very large. This is because; in many evaluations, only evaluation judgment is made by comparing the plan and the actual result, and check items about "when there is a significant difference between the plan and the actual result" and "when the evaluation judgment is made according to the increase or decrease of the output" are checked as "Not applicable".
	Third party QC > self-assessment	6 (3.7%)	
	Third party QC < self-assessment	2 (1.2%)	
	NA, with no items, with no answer	<b>103 (64.0%)</b>	
Sustainability (42 TC X 10 check items X + 5 TCDP X 10 check items + 5 GA X 10 check items = 520)	Third party QC = self-assessment	<b>366 (70.4%)</b>	● The results of self-assessment and third party QC are generally consistent. ● In many cases, third party QC scores higher than self-assessment.
	Third party QC > self-assessment	<b>86 (16.5%)</b>	
	Third party QC < self-assessment	35 (6.7%)	
	NA, with no items, with no answer	33 (6.3%)	

# 1. Overall trend

Evaluation criteria	Difference between self-assessment and third party QC	Number of items (rate to the population)	Comments
Conclusion / recommendation / lesson learned (42 TC X 4 check items X + 5 TCDP X 4 check items + 5 GA X 4 check items = 208)	Third party QC = self-assessment	124 (59.6%)	<ul style="list-style-type: none"> <li>In many cases, third party QC scores lower than the self-assessment. The reason is that the third party QC verifies the relationship between the recommendation / lesson learned and the contents in the evaluation report and the usefulness of the recommendation / lesson learned rather rigorously.</li> </ul>
	Third party QC > self-assessment	7 (3.4%)	
	Third party QC < self-assessment	19 (9.1%)	
	NA, with no items, with no answer	58 (27.9%)	
General items (42 TC X 3 check items X + 5 TCDP X 3 check items + 5 GA X 3 check items = 156)	Third party QC = self-assessment	70 (44.9%)	<ul style="list-style-type: none"> <li>There are many cases checked as "Not applicable". This is because many of the check items in "sample survey" are checked as "Not applicable".</li> <li>In many cases, third party QC scores higher than self-assessment.</li> </ul>
	Third party QC > self-assessment	12 (7.7%)	
	Third party QC < self-assessment	3 (1.9%)	
	NA, with no items, with no answer	71 (45.5%)	

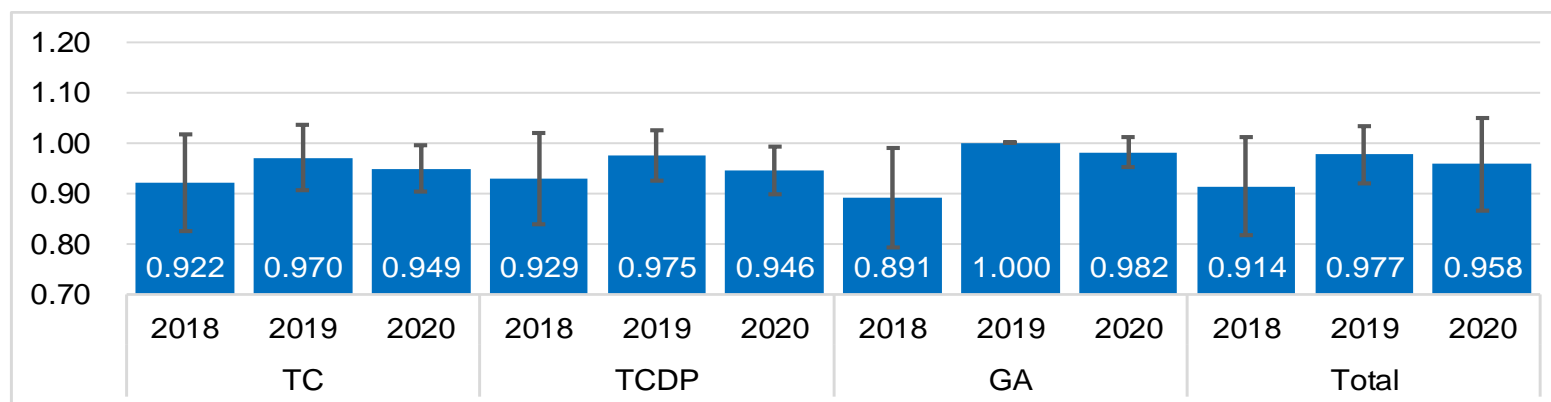


## 2. Analysis by evaluation criteria

## 2. Analysis by evaluation criteria (Relevance)

Average, dispersion, and standard deviation of standardized score of “Relevance” in 52 targeted projects

	TC			TCDP			GA			Total		
Year	2018	2019	2020	2018	2019	2020	2018	2019	2020	2018	2019	2020
Average	0.922	0.970	0.949	0.929	0.975	0.946	0.891	1.000	0.982	0.914	0.977	0.958
Dispersion	0.009	0.004	0.002	0.008	0.003	0.002	0.010	0.000	0.001	0.009	0.003	0.009
Standard deviation	0.096	0.065	0.046	0.091	0.050	0.048	0.099	0.000	0.031	0.097	0.056	0.093



Note: Bar graph: average of standardized score, error bar (error range):  $\pm$  standard deviation

- In 2020, standardized score of GA is higher than other schemes.
- Compared to FY2019, the standardized score has dropped slightly. There are some evaluations in which the explanation of the policy position of the project and the explanation of development needs were insufficient.
- Regarding the appropriateness of the approach, "not applicable" accounted for the majority because there are few evaluations in which "effectiveness / impact" and "sustainability" were judged to be "low".



## 2. Analysis by evaluation criteria (Relevance)

	Check item (TC)	Check item (TCDP)	Check item (GA)	Average (raw score)	Third party QC comment / analysis
1	[Development policy] Does the evaluation report mention (1) the names of the development policies of the country, (2) the years covered by these policies, and (3) clarification of the project's position in these policies at the time of ex-ante and ex-post evaluation, and (4) does the evaluation judgement align with them?		[Development policy] Does the evaluation report mention (1) the names of the development policies of the country, (2) the years covered by these policies, and (3) clarification of the project's position in these policies at the time of ex-ante and ex-post evaluation, and (4) does the evaluation judgement align with them?	TC: 1.833 TCDP: 2.000 GA: 2.000	<p>[Example of comments]</p> <ul style="list-style-type: none"> <li>✓ The policy at each time point is not described. (For phase integrated evaluation)</li> <li>✓ It is necessary to explain the policies of the target countries. (For wide area projects)</li> <li>✓ Target year unknown.</li> <li>✓ The target year does not match the project to be evaluated.</li> </ul> <p>[Analysis] In general, majority of the evaluations are without any problems.</p>
2	[Development needs] Does the evaluation report mention (1) the development needs of the country, (2) the information to indicate the existence of the needs (e.g., description of the context of the project, etc.) at the times of both ex-ante evaluation and project completion, and (3) does the evaluation judgement align with them?		[Development needs] Does the evaluation report mention (1) the development needs of the country, (2) the information to indicate the existence of the needs (e.g., description of the context of the project, etc.) at the times of both ex-ante evaluation and ex-post evaluation, and (3) does the evaluation judgement align with them?	TC: 1.881 TCDP: 2.000 GA: 2.000	<p>[Example of comments]</p> <ul style="list-style-type: none"> <li>✓ There is no explanation of the background or rationale for the needs (only the statement "it was thought necessary", etc.).</li> <li>✓ The needs at each time point are not described. (For phase integrated evaluation)</li> <li>✓ It is better to describe the development needs at the time of ex-ante evaluation and project completion (or ex-post evaluation). (e.g. presence / absence and changes of needs at each time point)</li> </ul> <p>[Analysis] There were some evaluation that are judged to be insufficiently explained in the background and grounds of the needs.</p>

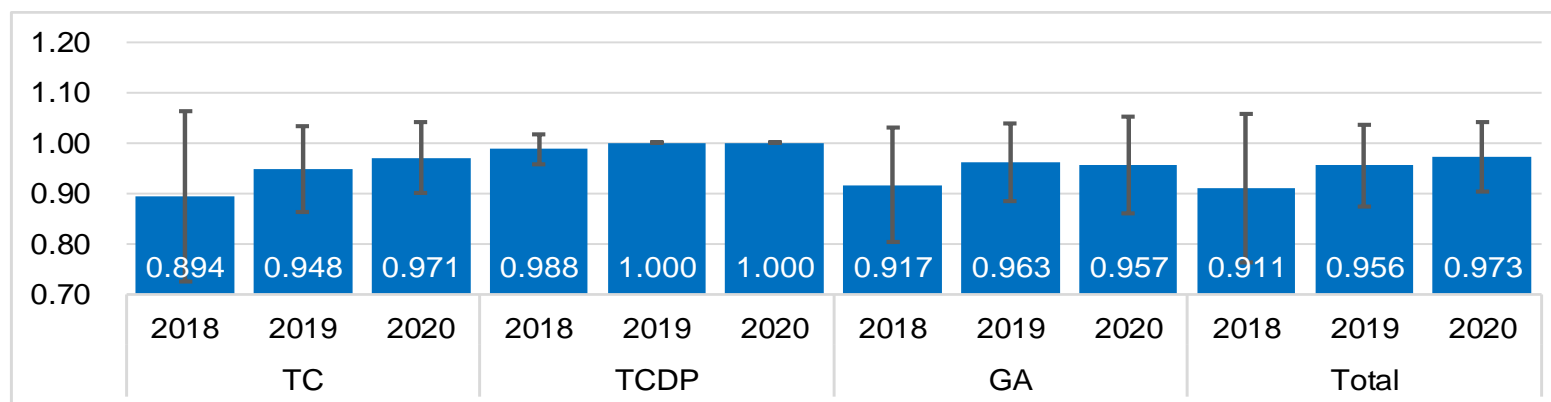
## 2. Analysis by evaluation criteria (Relevance)

	Check item (TC)	Check item (TCDP)	Check item (GA)	Average (raw score)	Third party QC comment / analysis
3	[Japan's ODA policy] Does the evaluation report mention (1) the names of the ODA policies for the country, (2) the years covered by these policies, and (3) relevant objectives, etc., at the time of ex-ante evaluation, and (4) does the evaluation judgement align with them?			TC: 1.905 TCDP: 2.000 GA: 2.000	<p>[Example of comments]</p> <ul style="list-style-type: none"> <li>✓ The target year does not match the project to be evaluated.</li> <li>✓ It would be better to have a detailed explanation of how the project was aligned with the priority issues.</li> </ul> <p>[Analysis]</p> <p>Many evaluations indicate only the statement that "XXX is important", and there are few evaluation that mentioned the policy position of the project.</p>
4	[Appropriateness] In case the project experienced a problem for Effectiveness/Impact and/or Sustainability, does the evaluation report reasonably analyze the appropriateness of the project based on the evidence?			TC: 2.000 TCDP: N/A GA: N/A	<p>[Example of comments]</p> <ul style="list-style-type: none"> <li>✓ The evaluation judgment of effectiveness / impact and sustainability is fair or higher, but the appropriateness of the approach is also analyzed. The description is appropriate.</li> </ul> <p>[Analysis]</p> <p>Most of the projects are scored as "not applicable" because few projects are evaluated as "low" in "effectiveness / impact" and "sustainability".</p>
5	[Evaluation Result] Is the evaluation judgment on Relevance as a whole consistent with the judgments on (1) the development policy, (2) the development needs and (3) Japan's ODA policy (and, if applicable, (4) appropriateness of the approach)?			TC: 1.976 TCDP: 2.000 GA: 2.000	<p>[Example of comments]</p> <ul style="list-style-type: none"> <li>✓ In the evaluation judgment in relevance, the verification of the appropriateness of the approach should also be taken into consideration. (in addition, it is necessary to explain the analysis process)</li> </ul> <p>[Analysis]</p> <p>Most of the evaluations are properly analyzed and described.</p>

## 2. Analysis by evaluation criteria (Effectiveness / impact)

Average, dispersion, and standard deviation of standardized score of “Effectiveness / impact” in 52 targeted projects

	TC			TCDP			GA			Total		
Year	2018	2019	2020	2018	2019	2020	2018	2019	2020	2018	2019	2020
Average	0.894	0.948	0.971	0.988	1.000	1.000	0.917	0.963	0.957	0.911	0.956	0.973
Dispersion	0.028	0.007	0.005	0.001	0.000	0.000	0.013	0.006	0.009	0.022	0.006	0.005
Standard deviation	0.169	0.086	0.070	0.029	0.000	0.000	0.114	0.077	0.096	0.148	0.081	0.069



Note: Bar graph: average of standardized score, error bar (error range):  $\pm$  standard deviation

- Overall, average score of 2020 has a slightly higher than the score of 2019.
- The average score of the TCDP tends to be higher than other schemes throughout the three years. In TCDP where master plans are often created, it seems that it is easy to clearly judge the achievement of outputs and outcomes and this characteristic may affect on the average score.

## 2. Analysis by evaluation criteria (Effectiveness / impact)

	Check item (TC)	Check item (TCDP)	Check item (GA)	Average (raw score)	Third party QC comment / analysis
6	<p>[Project Purpose] (1) Does the evaluation report present and properly analyze the indicator data through the time of project completion, and (2) are they consistent with the judgement of “achieved, partially achieved, or unachieved”? (3) If the indicator data cannot be presented, is an analysis based on data of reasonable supplementary information included?</p>	<p>[Achievement of the Objectives at the timing of project completion] Does the evaluation report present concrete evidence and properly analyze the status of delivered output and the achievement of the project purpose at the time of project completion?</p>	<p>[Effectiveness] With respect to the project objective(s), (1) does the evaluation report present and analyze the indicator data for the period between the project completion and the ex-post evaluation (including the target year)? (2) If the report fails to include certain indicator data, does it offer an analysis based on supplemental information designated in the EFM? (3) Does the report present the information on qualitative effects (including effects of the soft component if applicable)? (4) Does the evaluation judgement of Effectiveness consistent with them?</p>	<p>TC: 1.929 TCDP: 2.000 GA: 1.800</p>	<p>[Example of comments]</p> <ul style="list-style-type: none"> <li>✓ The specified project purpose / indicators do not match the data shown.</li> <li>✓ The data shown are inconsistent with the evaluation judgment.</li> <li>✓ It is necessary to explain a background for not achieving the target.</li> <li>✓ The data shown is insufficient to derive an evaluation judgment.</li> <li>✓ Initially, the target value has not been set, so it is difficult to judge whether it has been achieved or not.</li> <li>✓ Judgment seems to be appropriate considering the alternative data.</li> <li>✓ Some data is insufficient. (There are years when data is not shown, etc.)</li> <li>✓ Inappropriate indicators in view of project purpose. (e.g. it seems necessary to check the improvement or change of XXX rate, not the XXX rate itself, etc.)</li> </ul> <p>[Analysis] In most of the evaluation, the data description (considering the constraints of the evaluation survey) seemed appropriate. Some of the evaluation with unreasonable data and those with insufficient data were subject to deductions in third party QC score.</p>

## 2. Analysis by evaluation criteria (Effectiveness / impact)

	Check item (TC)	Check item (TCDP)	Check item (GA)	Average (raw score)	Third party QC comment / analysis
7	[Continuation Status of Project Effects] (1) Does the evaluation report present and properly analyze the data showing continued effects through the time of ex-post evaluation? (2) Does the judgement of "continued, partially continued or not continued" consistent with them? (3) If the report fails to include certain indicator data, does it offer an analysis based on supplemental information designated in the EFM?	No check item	[Effectiveness] With respect to the operation status of the facilities/equipment developed under the project and generation/continuation status of projects effects through the time of ex-post evaluation, (1) does the evaluation report present and analyze quantitative data such as capacity utilization, etc. or qualitative information? (2) Do they accordantly reflect on the evaluation judgement?	TC: 1.952 TCDP: No check item GA: 2.000	<p>[Example of comments]</p> <ul style="list-style-type: none"> <li>✓ Part of the data is insufficient (e.g. it is better to have data from the satisfaction survey not only at the time of project completion but also at the time of ex-post evaluation, as changes cannot be confirmed from the time of project completion to the time of ex-post evaluation, etc.)</li> <li>✓ The data presented is not sufficient to derive the evaluation judgement.</li> <li>✓ Although the data is insufficient, the evaluation judgment of "unverifiable" seems to be appropriate.</li> <li>✓ Evaluation is conducted using alternative data. Judgment is appropriate.</li> <li>✓ The data shown are inconsistent with the evaluation judgment.</li> </ul> <p>[Analysis] (Refer to the analysis column of check item 6)</p>

## 2. Analysis by evaluation criteria (Effectiveness / impact)

	Check item (TC)	Check item (TCDP)	Check item (GA)	Average (raw score)	Third party QC comment / analysis
8	<p>[Overall Goal] (1) Does the evaluation report present and correctly analyze the indicator data from the time of project completion to the time of ex-post evaluation including the target year? (2) Are the consistent with evaluation judgement? (3) If the report fails to include certain indicator data, does it offer an analysis based on supplemental information designated in the EFM?</p>	<p>[Utilization level of the proposed plan] (1) Does the evaluation report present and correctly analyze the indicator data showing the status of Utilization level about the proposed plan r? (2) If the report fails to include certain indicator data, does it offer an analysis based on supplemental information designated in the EFM?</p>	<p>[Impact] With respect to the expected impact (the part after "thereby contributing to" in "Objectives of the Project"), (1) does the evaluation report present and analyze quantitative data or qualitative information through the time of ex-post evaluation? (2) Do the accordantly reflect on the evaluation judgement?</p>	<p>TC: 1.929 TCDP: 2.000 GA: 2.000</p>	<p>[Example of comments]</p> <ul style="list-style-type: none"> <li>✓ Since the target is not indicated, it is difficult to judge whether the target had been achieved or not.</li> <li>✓ The data shown are inconsistent with the indicators.</li> <li>✓ The causal relationship between the data shown and the contribution of the project is weak or unknown.</li> <li>✓ Some data is missing.</li> <li>✓ The evaluation was conducted with alternative data, and the judgment seems to be appropriate.</li> <li>✓ The setting of alternative indicators is appropriate, but the data of the initially set indicators should also be shown and analyzed (if it is impossible, the reason should be stated).</li> <li>✓ The data shown are inconsistent with the evaluation judgment.</li> <li>✓ The evaluation judgment is a little questionable (e.g. it seems "not achieved" instead of "unverifiable").</li> <li>✓ Since only the data at the time of the ex-post evaluation is described, the improvement or change from the completion of the project is unknown.</li> <li>✓ It is difficult to understand the definition of the indicator and the path of the project effect.</li> </ul> <p>[Analysis] (Refer to the analysis column of check item 6)</p>

## 2. Analysis by evaluation criteria (Effectiveness / impact)

	Check item (TC)	Check item (TCDP)	Check item (GA)	Average (raw score)	Third party QC comment / analysis
9	[Project Purpose & Overall Goal] In cases as there are problems in indicators to be verified (such as no target value, inappropriate indicator for continuation status of project effects, vague definition of indicator ), is the project judged "Not verified", judged based on supplemental information or judged with reasonable explanations?	No check item		TC: 1.821 TCDP: No check item GA: No check item	<p>[Example of comments]</p> <ul style="list-style-type: none"> <li>✓ The alternative indicator is used to determine the achievement of the project purpose, but it is necessary to explain the reason of this.</li> <li>✓ The initial plan does not provide a target, but the evaluation provides sufficient data to derive the evaluation judgment.</li> <li>✓ Although no overall goal had been set, the overall goal assumed in the ex-post evaluation and evaluation judgments seems appropriate. (For SATREPS project)</li> <li>✓ The data and evaluation judgments described are generally appropriate, taking into account the restrictions of the evaluation survey.</li> <li>✓ The evaluation judgment of "unverifiable" is appropriate.</li> <li>✓ It was difficult to understand the definitions of project purpose and overall goal, and each indicator seems to point to almost the same content.</li> <li>✓ Inappropriate indicator setting.</li> <li>✓ Not enough data is shown to make an evaluation judgement.</li> </ul> <p>[Analysis] There were many evaluations those are scored as "not applicable", but the applicable evaluation reports are generally described appropriately.</p>

## 2. Analysis by evaluation criteria (Effectiveness / impact)

	Check item (TC)	Check item (TCDP)	Check item (GA)	Average (raw score)	Third party QC comment / analysis
10	(Other positive and negative impacts) If there are other positive and negative impacts, are they correctly presented and analyzed?			TC: 2.000 TCDP: 2.000 GA: 2.000	[Analysis] There were many evaluations those are scored as "not applicable". There are no particular comments. They are generally properly described.
11	Are the grounds for judging individual sub-items of Evaluation/Impact objective and neutral (i.e., Is the judgment not made only based on an opinion of one person or impression? Is the objective not judged as achieved without reasonable explanations even though the indicator values substantially fall short of target values? Is the objective not judged as achieved even though the target values or actual values are unavailable? No it isn't => tick "Yes"; Yes it is => tick "No")			TC: 2.000 TCDP: 2.000 GA: 1.800	[Example of comments] ✓ The data shown are inconsistent with the evaluation judgment. ✓ Evaluation judgment is a little unreasonable. (e.g. it seems to be "not achieved", not "unverifiable".)  [Analysis] Most of the evaluations are generally properly described.
12	In case that the factors other than the project mainly affected achievement/non-achievement on individual sub-items and for the increase/decrease in indicator values, does the evaluation report state reasons for it?			TC: 2.000 TCDP: 2.000 GA: 2.000	[Example of comments] ✓ XXXX is mentioned as an external factor. ✓ External factors are taken into consideration in the evaluation judgment, and the judgment is rational.  [Analysis] There are many evaluations those do not describe external factors (it is difficult for third party QC to confirm whether or not there are external factors), and there are many checks for "not applicable". The evaluations those describe the external factors are generally appropriate.



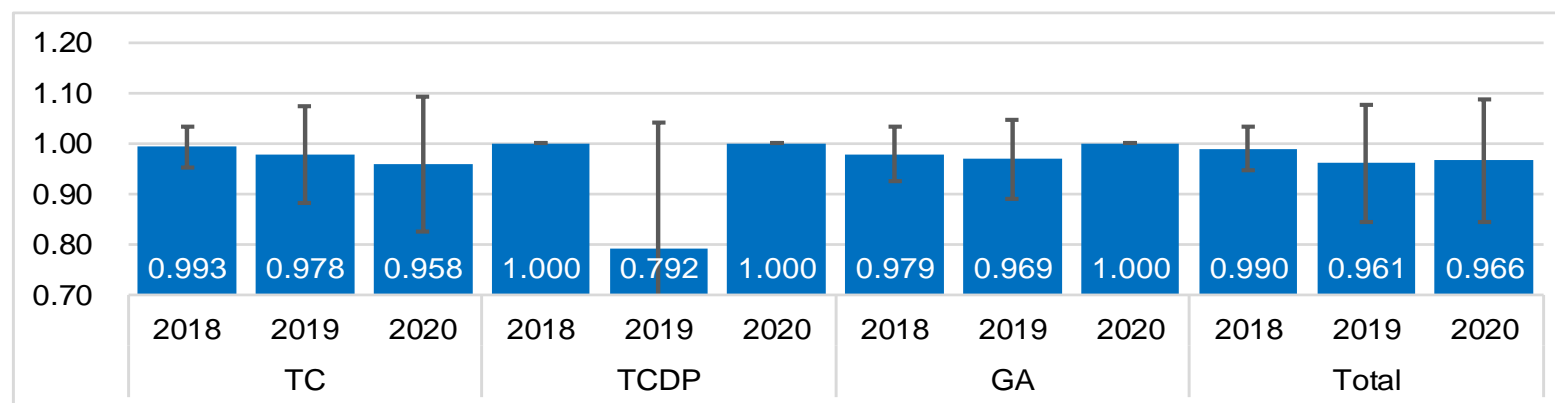
## 2. Analysis by evaluation criteria (Effectiveness / impact)

	Check item (TC)	Check item (TCDP)	Check item (GA)	Average (raw score)	Third party QC comment / analysis
13	<p>[Evaluation Result] Is the evaluation judgment on Effectiveness/Impact as a whole consistent with the judgments on 1. the degree to which Project Purpose has been achieved, 2. the status of the continued effects of Project Purpose toward achieving Overall Goal, and, 3. the degree to which Overall Goal has been achieved? 4. If other positive and negative impacts are stated and the impact is considered to be large, is the evaluation judgment reasonably taken into consideration?</p>	<p>[Evaluation Result] Is the evaluation judgement on Effectiveness/Impact as a whole consistent with the judgements on 1. the degree of the achievement of the objectives by the project completion, and 2. the status of utilization of the proposed plan by the time of ex-post evaluation. 3. If other positive and negative impacts are stated and the impact is considered to be large, is the evaluation judgment reasonably taken into consideration?</p>	<p>[Evaluation Result] Is the evaluation judgment on Effectiveness/Impact as a whole consistent with the judgments on 1. the quantitative and qualitative effects (Effectiveness) and 2. the expected impact. 3. If other positive/negative impacts are stated and they are considered significant, does the evaluation judgment reflect them?</p>	<p>TC: 1.952 TCDP: 2.000 GA: 1.800</p>	<p>[Example of comments]</p> <ul style="list-style-type: none"> <li>✓ Evaluation judgment seems to be not rational. (e.g. in view of the contents of the description, it cannot be judged that the overall goal had been achieved, and it seems that it is not rational to judge that the overall effectiveness and impact are "high". It is necessary to explain the reason for the judgment.)</li> <li>✓ The definition of project purpose and overall goals is ambiguous. Evaluation judgment is influenced by how these are perceived. Therefore, it is difficult to confirm its appropriateness with a third party QC.</li> </ul> <p>[Analysis] Many made evaluation judgments that seems appropriate. From the perspective of third party QC, projects those seem unreasonable in their judgement in effectiveness / impact were subject to deductions in score.</p>

## 2. Analysis by evaluation criteria (Efficiency)

Average, dispersion, and standard deviation of standardized score of “Efficiency” in 52 targeted projects

	TC			TCDP			GA			Total		
Year	2018	2019	2020	2018	2019	2020	2018	2019	2020	2018	2019	2020
Average	0.993	0.978	0.958	1.000	0.792	1.000	0.979	0.969	1.000	0.990	0.961	0.966
Dispersion	0.002	0.009	0.018	0.000	0.063	0.000	0.003	0.006	0.000	0.002	0.014	0.015
Standard deviation	0.041	0.095	0.134	0.000	0.250	0.000	0.055	0.078	0.000	0.044	0.117	0.122



Note: Bar graph: average of standardized score, error bar (error range):  $\pm$  standard deviation

- In 2020, the average score for TCDP and GA is high. The number of projects is small in both schemes, and it seems that the results of third party QC for individual projects have an effect on average score.
- The dispersion is large compared to other evaluation items. There are many projects / check items those are scored as “not applicable”, and if they are subject to scoring, there will be a clear difference in scoring, and as a result, the dispersion would become large.

## 2. Analysis by evaluation criteria (Efficiency)

	Check item (TC)	Check item (TCDP)	Check item (GA)	Average (raw score)	Third party QC comment / analysis
14	No check item		In case there is a significant gap between the planned and actual outputs, does the evaluation report mention it along with reason(s) in "Efficiency" or other sections (such as "Project Outline" and "Effectiveness/Impact")?	TC: No check item TCDP: No check item GA: 2.000	[Analysis] There is no particular comment. The check item is only for GA, and there are many checks for "not applicable".
15	In case there was a significant discrepancy between the planned and actual inputs (project period and project cost), (1) does the evaluation report mention the reason(s), and (2) is it rational?			TC: 1.333 TCDP: 2.000 GA: 2.000	[Example of comments] ✓ The actual period / cost is less than 150% of the plan, but the reason for the difference is stated. The reason is reasonable. ✓ The reason should be explained because the cost of the TC is more than 150% of the plan. ✓ There was the increase in output, but it is better to explain the content of the increased output a little more in detail in terms of effectiveness / impact, or efficiency. [Analysis] There are many checks in "not applicable". Most of the TC are "not applicable", but as a result, the average score of the raw score become low because the scored evaluations are often deducted in the score. Although the reason is stated, there are a few evaluations those seem necessary to be explained in a little more in detail.
16	If the assessment of the inputs is made based on the increase or decrease in the outputs rather than on a simple comparison between the planned and actual inputs, (1) does the evaluation report explain it in relation to the outcomes? (2) Is the explanation reasonable?			TC: 1.500 TCDP: 2.000 GA: 2.000	[Example of comments] ✓ Regarding the statement that "some outputs were partial produced", it should also be explained which output was partial and how. [Analysis] There are many evaluations checked in "not applicable". Most of TC are "not applicable", but as a result, the average score of the raw score was low because the scored evaluations are often deducted in score.

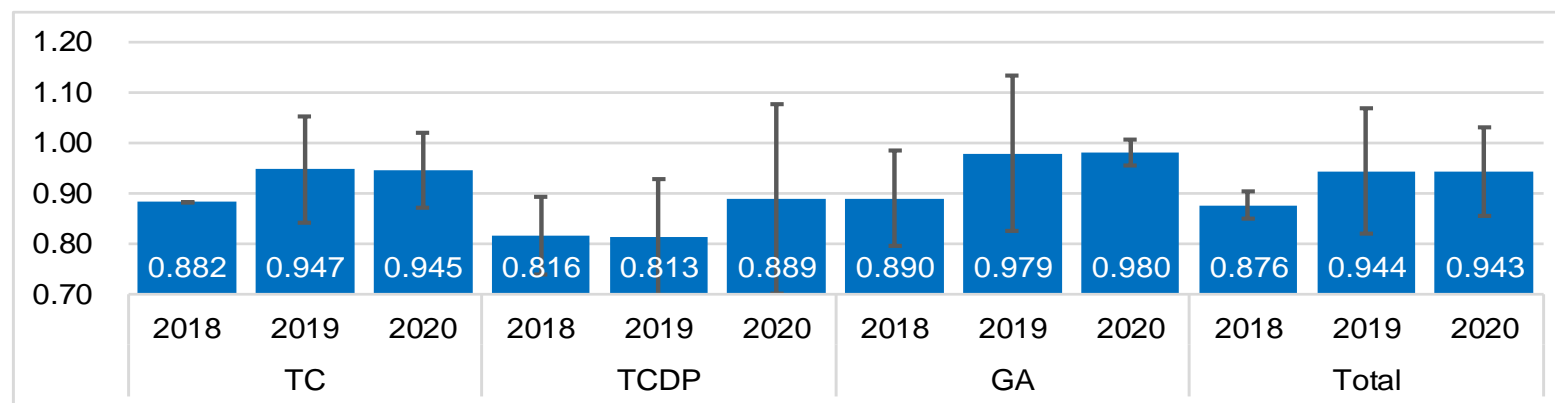
## 2. Analysis by evaluation criteria (Efficiency)

	Check item (TC)	Check item (TCDP)	Check item (GA)	Average (raw score)	Third party QC comment / analysis
17	[Evaluation Result] Is the evaluation judgment on Efficiency as a whole consistent with the judgments on (1) the project period and (2) the project cost?			TC: 1.952 TCDP: 2.000 GA: 2.000	<p>[Example of comments]</p> <ul style="list-style-type: none"> <li>✓ The reason why the actual period exceeded the planned period is reasonable. However, when making an evaluation judgement considering the exclusion of the period when the project was interrupted due to a natural disaster or deterioration of public security, after explaining that in the evaluation (how much it is excluded, etc.), the ratio to the plan should be stated. If the ratio that reflects the above is not stated, the appropriateness of the evaluation judgment cannot be judged.</li> <li>✓ (In the case where the initial planned period is 18 months and was extended to 26 months by additional components, and the planned period is evaluated as 26 months) The planned period should be set to the initial 18 months, and the evaluation judgement should be made considering the increase in output.</li> </ul> <p>[Analysis] Generally, most of the evaluation do not have problem and are properly described.</p>

## 2. Analysis by evaluation criteria (Sustainability)

Average, dispersion, and standard deviation of standardized score of “Sustainability” in 52 targeted projects

	TC			TCDP			GA			Total		
Year	2018	2019	2020	2018	2019	2020	2018	2019	2020	2018	2019	2020
Average	0.882	0.947	0.945	0.816	0.813	0.889	0.890	0.979	0.980	0.876	0.944	0.943
Dispersion	0.011	0.006	0.066	0.013	0.036	0.009	0.024	0.001	0.001	0.015	0.008	0.006
Standard deviation	0.105	0.075	0.077	0.114	0.189	0.095	0.155	0.026	0.027	0.124	0.089	0.077



Note: Bar graph: average of standardized score, error bar (error range):  $\pm$  standard deviation

- In 2020, the average points of standardized score for each scheme are almost the same as in 2019. Throughout the three years, TCDP tends to have a slightly lower standardized score than other schemes.
- Third party QC check items are set for each aspect of sustainability (policy, organization, technics, and finance). Although there are some evaluations those seem to be insufficiently described, the quality of the description is generally ensured and the standardized score is high. The only check item related to "factors and prospects" related to sustainability issues is that the description of the prospects is insufficient in many evaluations, and the score is deducted.

## 2. Analysis by evaluation criteria (Sustainability)

	Check item (TC)	Check item (TCDP)	Check item (GA)	Average (raw score)	Third party QC comment / analysis
18	[Policy Aspect] Does the evaluation report mention (1) the names of the development policies of the country, (2) the target years of the policies, and (3) relevant objectives, etc., at the time of ex-post evaluation? (4) Are they consistent with evaluation judgement?		No check item	TC: 1.905 TCDP: 2.000 GA: No check item	<p>[Example of comments]</p> <ul style="list-style-type: none"> <li>✓ The target year of the policy is not stated.</li> <li>✓ It is necessary to explain how the policy is positioned in analyzing the sustainability of the project in terms of policy.</li> <li>✓ It is better if there is an explanation of the contents of the decree.</li> </ul> <p>[Analysis]</p> <p>Although there are some evaluations those are incomplete in format and lacked information, most of the evaluations are not problematic in general.</p>
19	[Institutional Aspect] Does the evaluation report mention (1) the organization(s) responsible for sustaining the project effects, and (2) make the roles of such organization(s) in sustaining the project effects clear?	[Institutional Aspect] Does the evaluation report mention (1) the organization(s) responsible for promoting the proposed development plan, and (2) make the roles of such organization(s) in promoting the proposed development plan clear?	[Institutional Aspect] Does the evaluation report mention (1) the organization(s) responsible for sustaining the project effects, and (2) make the roles of such organization(s) in sustaining the project effects clear?	TC: 1.905 TCDP: 1.400 GA: 2.000	<p>[Example of comments]</p> <ul style="list-style-type: none"> <li>✓ The role of each institution is difficult to understand.</li> </ul> <p>[Analysis]</p> <p>The average score of TCDP is a little low, but it is rather influenced by the third party QC results for individual projects than the characteristics of the scheme. On the other hand, with regard to GA, the average score seems to be high because the role of the institution is clearer than that of the TC / TCDP as an institution responsible for the maintenance of the provided materials and equipment.</p> <p>Insufficient explanation is confirmed in some of the evaluations, and there are differences in the amount of information and the way of thinking of evaluation judgment for each project, but generally there are few evaluations those have problems in total.</p>

## 2. Analysis by evaluation criteria (Sustainability)

	Check item (TC)	Check item (TCDP)	Check item (GA)	Average (raw score)	Third party QC comment / analysis
20	[Institutional Aspect] Does the evaluation report (1) make it clear the reported organizational structure is necessary for what specific activities/services (e.g., dissemination of the model) ? (2) Does the evaluation report present the data about the number of staff, and (3) the opinion of concerned party?	[Institutional Aspect] Does the evaluation report (1) make it clear the reported organizational structure is necessary for what specific activities to promote the proposed development plan? (2) Does the evaluation report present the data about the number of staff, and (3) the opinion of concerned party?	[Institutional Aspect] Does the evaluation report present (1) the data of the organizational structure and the number of staff, and (2) opinion of concerned party?	TC: 1.857 TCDP: 1.600 GA: 2.000	<p>[Example of comments]</p> <ul style="list-style-type: none"> <li>✓ Insufficient explanation of the institution.</li> <li>✓ Data on organizational structure and personnel are not shown, and the basis for evaluation judgment is difficult to understand.</li> <li>✓ It is difficult to understand what the necessary system is for.</li> <li>✓ The evaluation judgment is made only from the viewpoint of CP, and the basis of the judgement is a little weak.</li> </ul> <p>[Analysis] (Refer to the analysis column of check item 19)</p>
21	[Institutional Aspect] With respect to the grounds on which the institutional levels such as the organizational structure and the number of staff are judged to be sufficient/insufficient, (1) does the evaluation report show data in the explanation? (2) Is the judgement consistent with analysis?			TC: 1.905 TCDP: 1.600 GA: 2.000	<p>[Example of comments]</p> <ul style="list-style-type: none"> <li>✓ The rationale for why the institution can be judged to be sustainable is not stated.</li> <li>✓ There are few explanations, but the evaluation judgment that the prospect is unclear is appropriate.</li> <li>✓ It seems that the evaluation judgment is derived only from the hearing results, and there are few explanations on the ground of judgement.</li> <li>✓ The basis for the evaluation judgment is rather weak, but it seems unavoidable due to the restrictions of the evaluation survey.</li> <li>✓ In the conclusion part, it is said that there is a problem in the institutional aspect, but it is difficult to understand the specific problem.</li> </ul> <p>[Analysis] (Refer to the analysis column of check item 19)</p>

## 2. Analysis by evaluation criteria (Sustainability)

	Check item (TC)	Check item (TCDP)	Check item (GA)	Average (raw score)	Third party QC comment / analysis
22	<p>[Technical Aspect] Does the evaluation report (1) make it clear the reported technique is necessary for what specific activity/service (e.g., dissemination of the model)? (2) Does the evaluation report present the data for the grounds on which it judges the technical level to be sufficient/insufficient, and (3) opinion of concerned party??</p>	<p>[Technical Aspect] Does the evaluation report show (1) the data of technique, and (2) make it clear the reported technique is necessary for what specific activity to promote the proposed development plan? Does the evaluation report present the data for the grounds on which it judges the technical level to be sufficient/insufficient, and opinion of concerned party?</p>	<p>[Technical Aspect] Does the evaluation report present (1) the data for the grounds on which it judges the technical level to be sufficient/insufficient and (2) opinion of concerned party?</p>	<p>TC: 1.976 TCDP: 1.800 GA: 2.000</p>	<p>[Example of comments] ✓ No data on the techniques is shown.</p> <p>[Analysis] The average score of TCDP is a little low, but this is more influenced by the third party QC results for individual projects than the characteristics of the scheme. On the other hand, GA is mainly evaluated from the viewpoint of the technical ability of equipment maintenance, and it can be said that the status of equipment maintenance indicates the technical sustainability. It seems that the viewpoint of evaluation is simple and the score was higher than that of TC and TCDP. Insufficient explanation is confirmed in some of the evaluations, and there are differences in the amount of information and the way of thinking for project, but most of the evaluations are described properly in total. In particular, regarding the technical sustainability of TC and TCDP, it is assumed that it would be difficult to provide a clear basis for judgment, and in some cases, judgment is made only from the results of hearings from CP. However, the third party QC understand that it is unavoidable due to the restrictions of the evaluation survey.</p>



## 2. Analysis by evaluation criteria (Sustainability)

	Check item (TC)	Check item (TCDP)	Check item (GA)	Average (raw score)	Third party QC comment / analysis
23	[Technical Aspect] With respect to the grounds on which the technical levels are judged to be sufficient/insufficient, (1) does the evaluation report show data in the explanation? (2) Is the judgement consistent with analysis?			TC: 1.952 TCDP: 1.600 GA: 2.000	<p>[Example of comments]</p> <ul style="list-style-type: none"> <li>✓ The basis for the evaluation judgment is rather weak, but it seems unavoidable due to the restrictions of the evaluation survey.</li> <li>✓ It is stated that there are some issues in technical sustainability, but the issues cannot be read from the description in the evaluation.</li> <li>✓ It is appropriate because the evaluation judgment is derived by taking into account the fact that the data is insufficient.</li> <li>✓ It is said that technical sustainability needs improvement, but it is difficult to understand what and why improvement is needed.</li> <li>✓ Insufficient financial resources have been shown as an issue, but insufficient financial resources should be taken into account in the evaluation judgment of financial aspects, not technical aspects.</li> <li>✓ (Regarding the evaluation that states there are some issues regarding the utilization status of the developed teaching materials) This kind of information should be considered in the evaluation of the continuation status of the effect in the effectiveness.</li> </ul> <p>[Analysis] (Refer to the analysis column of check item 22)</p>
24	[Financial Aspect] Does the evaluation report present (1) the data regarding the comparison between the planned and actual budget, future budget, and (2) the opinion of concerned party? (3) If the report fails to present and analyze the budget data, is the financial condition not judged as "no problem," etc., without presenting supplemental information? (No it isn't => tick "Yes"; Yes it is => tick "No")			TC: 1.929 TCDP: 2.000 GA: 2.000	<p>[Example of comments]</p> <ul style="list-style-type: none"> <li>✓ No budget data is presented.</li> <li>✓ It would be better to have a detailed explanation such as the presentation of specific budget data, not only to state that the budget for the proposed project be secured. (for the evaluation of TCDP)</li> </ul> <p>[Analysis] Even if the figures in the financial data is stated, there are few projects that can directly judge the financial sustainability, and in many evaluations, the judgment is made from qualitative information such as interviews with the implementing agency. There are many evaluations that did not include financial data, but it seems unavoidable due to the limitation of the evaluation survey.</p>

## 2. Analysis by evaluation criteria (Sustainability)

	Check item (TC)	Check item (TCDP)	Check item (GA)	Average (raw score)	Third party QC comment / analysis
25	[Financial Aspect] With respect to the grounds on which the financial condition is judged to be sufficient/insufficient, (1) does the evaluation report show data in the explanation? (2) Is judgement consistent with the analysis?			TC: 1.929 TCDP: 2.000 GA: 2.000	<p>[Example of comments]</p> <ul style="list-style-type: none"> <li>✓ If securing funds for activities is a financial issue, it is better to describe it so that it can be understood.</li> <li>✓ There is a description of budget data, but it is difficult to understand the basis why it can be said to be sufficient.</li> <li>✓ The data is limited and the basis for evaluation judgment is rather weak, but it seems unavoidable due to the restrictions of the evaluation survey.</li> <li>✓ It seems appropriate because the evaluation judgment is derived by taking into account the fact that the data is insufficient.</li> <li>✓ There are few explanations, but the evaluation judgment that the prospect is unclear is appropriate.</li> </ul> <p>[Analysis] (Refer to the analysis column of check item 24)</p>
26	If issues are found in individual aspects, does the evaluation report mention (1) the factors behind such issues and (2) prospects?			TC: 1.486 TCDP: 1.750 GA: 1.500	<p>[Example of comments]</p> <ul style="list-style-type: none"> <li>✓ There is a description of the current status of the issues, but there are few explanations of the factors.</li> <li>✓ There is a description of the current status of the issues, but there are few explanations of the prospects.</li> <li>✓ Since the issue is unknown (for the evaluation for which no specific issue is shown), third party QC considers it as "not applicable".</li> </ul> <p>[Analysis] It is the check item with the many evaluations deducted in score. In particular, there are many evaluations for which no description is found regarding the prospects. The average raw score is low for all three schemes.</p>

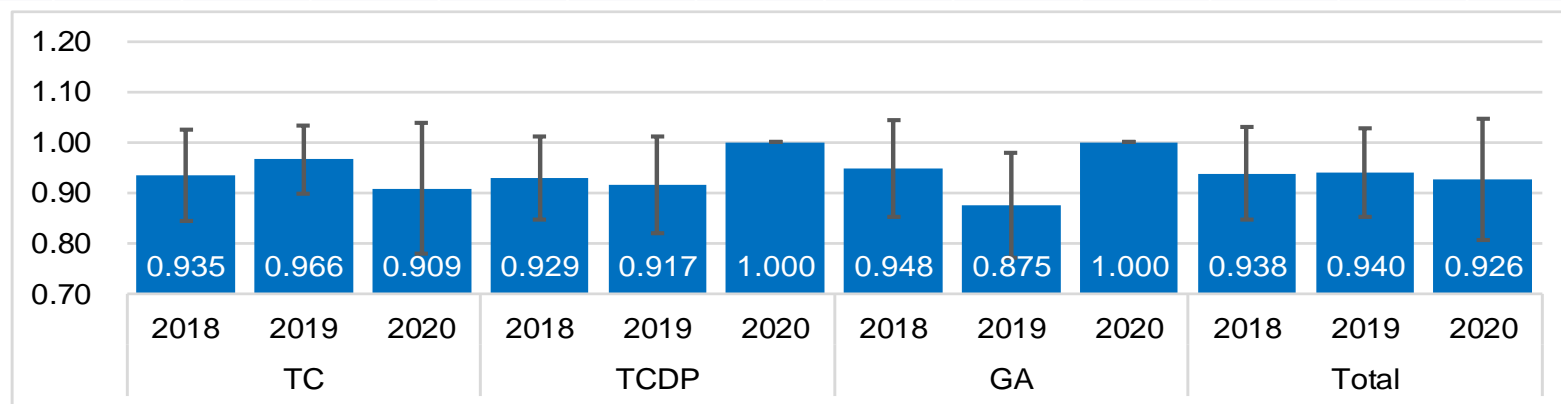
## 2. Analysis by evaluation criteria (Sustainability)

	Check item (TC)	Check item (TCDP)	Check item (GA)	Average (raw score)	Third party QC comment / analysis
27	[Evaluation Result] Is the evaluation judgment on Sustainability as a whole consistent with the judgments on (1) policy aspect, (2) institutional aspect, (3) technical aspect and (4) financial aspect?		[Evaluation Result] Is the evaluation judgment on Sustainability as a whole consistent with the judgments on (1) institutional aspect, (2) technical aspect and (3) financial aspect?	TC: 2.000 TCDP: 2.000 GA: 2.000	<p>[Example of comments]</p> <p>✓ Questions about the content of the description (e.g. it seems that there are some issues not only on the XXX aspect but also on the XXX aspect, though the evaluation judgment of the overall sustainability is appropriate.)</p> <p>[Analysis]</p> <p>There are no projects in which there is a contradiction in the evaluation judgment of the overall sustainability from the judgment of each aspect.</p>
28	No check item		(1) Does the evaluation report mention the situation about Operation and Maintenance of facilities and equipment on institutional aspect, technical aspect, or financial aspect properly? (2) Does this reflect on evaluation judgement?	TC: No check item TCDP: No check item GA: 2.000	<p>[Analysis]</p> <p>There is no particular comment. In all GA projects, the operation and maintenance status of equipment and materials are properly described.</p>

## 2. Analysis by evaluation criteria (Conclusion / Recommendation / lesson learned)

Average, dispersion, and standard deviation of standardized score of “Conclusion / Recommendation / lesson learned” in 52 targeted projects

	TC			TCDP			GA			Total		
Year	2018	2019	2020	2018	2019	2020	2018	2019	2020	2018	2019	2020
Average	0.935	0.966	0.909	0.929	0.917	1.000	0.948	0.875	1.000	0.938	0.940	0.926
Dispersion	0.008	0.005	0.017	0.007	0.009	0.000	0.009	0.011	0.000	0.008	0.008	0.015
Standard deviation	0.090	0.068	0.129	0.082	0.096	0.000	0.097	0.104	0.000	0.092	0.088	0.121



Note: Bar graph: average of standardized score, error bar (error range):  $\pm$  standard deviation

- Regarding the conclusion, all the projects derived overall evaluation judgement that is consistent with the judgment of each evaluation item, and no problematic description are found.
- Regarding recommendations and lessons learned, the average score is slightly lower. There are some recommendations and lessons learned that are not closely related to the contents in the evaluation report. From the viewpoint of usefulness and concreteness, there is also an evaluation in which the score by the third party QC is deducted. Since the recommendations / lessons learned are stated in most of the evaluations, most of the check items of "Is it appropriate if there are no recommendations / lessons learned?" are checked as "not applicable".

## 2. Analysis by evaluation criteria (Conclusion / Recommendation / lesson learned)

	Check item (TC)	Check item (TCDP)	Check item (GA)	Average (raw score)	Third party QC comment / analysis
29	[Summary of the Evaluation] Is the overall evaluation judgment consistent with the description of evaluation results?			TC: 2.000 TCDP: 2.000 GA: 2.000	[Analysis] There is no particular comment. In all the evaluations, an overall evaluation judgement is consistent with the judgment of each evaluation item.
30	[Recommendations] (1) Is the reason(s) why the proposed recommendation should be followed (i.e., issues, etc.) in relevant sections of evaluation report ? (2) Is the necessity of recommendation explained enough?			TC: 1.675 TCDP: 2.000 GA: 2.000	<p>[Example of comments]</p> <ul style="list-style-type: none"> <li>✓ The contents in the evaluation report and the recommendations are inconsistent. (e.g. while the background of project states the lack of CP ability, the recommendation cites that the CP ability is originally high as a factor for the success of the project.)</li> <li>✓ It is necessary to explain not only the content of the recommendation but also the necessity of the recommendation.</li> <li>✓ The background for which recommendations are needed is not mentioned in the evaluation report.</li> <li>✓ It is difficult to understand the relation between the contents in the evaluation report and the recommendation.</li> <li>✓ It is too general.</li> <li>✓ It is doubtful that the problem will be solved only by the actions stated in recommendation. It is better to have a little more detailed and concrete description.</li> <li>✓ The recommendation is lacking in concreteness.</li> <li>✓ If the recommendation help promote effectiveness, it should better to be mentioned in the effectiveness / impact and sustainability parts.</li> </ul> <p>[Analysis] The average raw score of TC is low. Although the average raw score for TCDP and GA is high, it is more influenced by the results of third party QC for individual projects rather than the characteristics of the scheme. There are many evaluations in which recommendations are derived that are not closely related to the contents in the evaluation report. From the viewpoint of usefulness and concreteness, some projects are deducted in score by third party QC.</p>

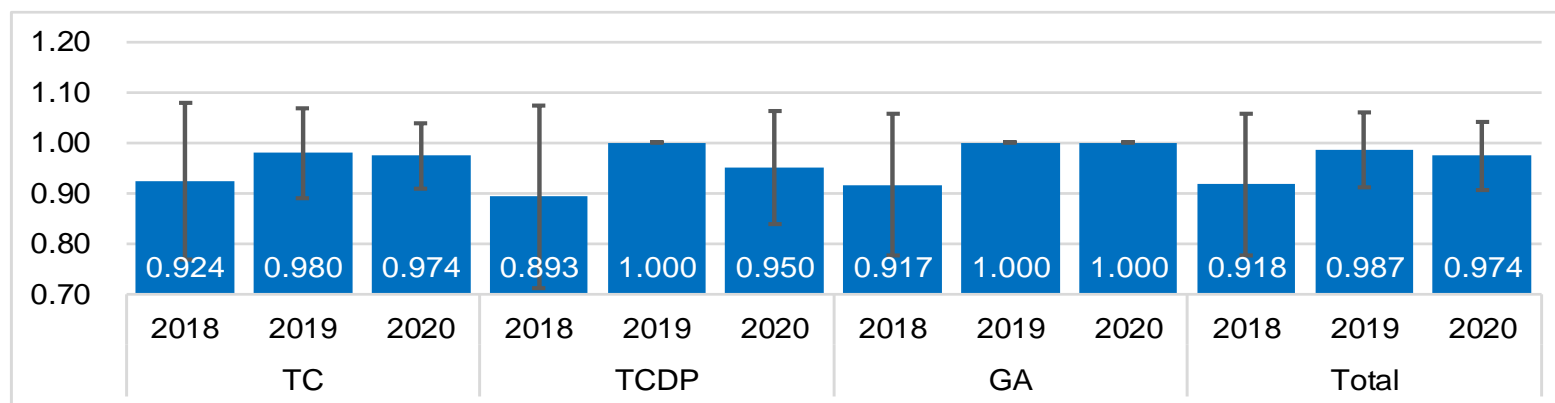
## 2. Analysis by evaluation criteria (Conclusion / Recommendation / lesson learned)

	Check item (TC)	Check item (TCDP)	Check item (GA)	Average (raw score)	Third party QC comment / analysis
31	[Lessons Learned] (1) Does the evaluation report mention in relevant sections the finding from which you drew the lessons learned? (2) Does this bring the useful lessons learned?			TC: 1.750 TCDP: 2.000 GA: 2.000	<p>[Example of comments]</p> <ul style="list-style-type: none"> <li>✓ It would be better to explain the necessity and usefulness of the lesson learned in a little more in detail.</li> <li>✓ It is necessary to explain not only the content of the lessons learned, but also the background of the lessons learned.</li> <li>✓ It is too general.</li> <li>✓ The lessons learned from changing the approach in light of external factors could also be mentioned in the effectiveness / impact parts.</li> <li>✓ (Regarding the lessons learned about the success factors of the project) It would be nice to explain in a little more detail why the project activities worked effectively.</li> <li>✓ The background of the issues shown in the lessons is not clear in the evaluation report.</li> </ul> <p>[Analysis]</p> <p>The average raw score of TC is low. Although the average raw score for TCDP and GA is high, it is more influenced by the results of third party QC for individual projects rather than the characteristics of the scheme.</p> <p>There are many evaluations in which lessons learned are derived that are not closely related to the contents in the evaluation report. From the viewpoint of usefulness and concreteness, some projects are deducted in score by third party QC.</p>
32	In case of no description of recommendation or lessons learned, is such way of description adequate? (Example of inadequate description: There is significant problem in impact or sustainability, but there is no recommendation for coping with the problem; There is factor analysis that might lead recommendation, but there is no recommendation derived, etc.)			TC: 2.000 TCDP: N/A GA: 2.000	<p>[Example of comments]</p> <ul style="list-style-type: none"> <li>✓ There is no description of lessons learned, but whether it is appropriate or not cannot be judged from the description of the evaluation itself.</li> </ul> <p>[Analysis]</p> <p>There is no particular comment. Most of the evaluations contains recommendations and lessons learned, so most of the evaluations are checked as "not applicable".</p>

## 2. Analysis by evaluation criteria (General items)

Average, dispersion, and standard deviation of standardized score of “General items” in 52 targeted projects

	TC			TCDP			GA			Total		
Year	2018	2019	2020	2018	2019	2020	2018	2019	2020	2018	2019	2020
Average	0.924	0.980	0.974	0.893	1.000	0.950	0.917	1.000	1.000	0.918	0.987	0.974
Dispersion	0.024	0.008	0.005	0.033	0.000	0.013	0.020	0.000	0.000	0.024	0.005	0.005
Standard deviation	0.155	0.090	0.065	0.182	0.000	0.112	0.141	0.000	0.000	0.141	0.074	0.067



Note: Bar graph: average of standardized score, error bar (error range):  $\pm$  standard deviation

- In 2020, most of the evaluations are properly described in general, and the standardized score is high.
- However, as for the third party QC check items that correspond to the sample survey, there are very few evaluations for which the sample survey was conducted, so most of them are checked as "not applicable". The average raw score of the check item is slightly lower, partly because the number of evaluations to be scored is small.

## 2. Analysis by evaluation criteria (General items)

	Check item (TC)	Check item (TCDP)	Check item (GA)	Average (raw score)	Third party QC comment / analysis
33	If there are important constraints on evaluation (e.g., inappropriate indicator(s), inadequate information collection, etc.), (1) does the evaluation report explain them and (2) how they might have affected evaluation judgment?			TC: 2.000 TCDP: 2.000 GA: 2.000	[Analysis] There are no particular comments. It seems that the evaluations, which contained restrictions and points to be noted, are generally properly described. However, it is difficult for third party QC to judge whether there are any restrictions on evaluation or points to be noted, and all the evaluations that are without the description are checked as "Not applicable".
34	Does the evaluation report identify the sources for all information it presents? ("According to the implementing agency" is acceptable)			TC: 2.000 TCDP: 1.800 GA: 2.000	[Example of comments] ✓ There is information whose source is difficult to understand for the data indicated in effectiveness / impact. [Analysis] Sources are indicated for almost all evaluations. However, it was also difficult to judge whether or not "all judgment grounds" are stated by the third party QC. (As for the way of describing the information source, it seems that it will be clearer if the rules of description of sources are clearly defined in the internal evaluation manual, etc.)
35	In case the evaluation report states it conducted a sample survey, (1) does it mention the number of valid responses and (2) the attributes of the respondents?			TC: 1.250 TCDP:N/A GA: N/A	[Example of comments] ✓ It is better to describe the attributes of the sample and the selected method. ✓ There is no description about the exact number of samples and attributes. [Analysis] There are few evaluations for which sample surveys were conducted, and the number of "not applicable" cases increases. Since it is not stated in the evaluation that it is a sample survey, the third party QC will confirm whether or not the sample survey is carried out from the contents of the evaluation report. In many cases, it is difficult to judge whether it is a sample survey or just an interview with the people concerned. (Regarding the description way for the sample survey, it seems that it will be clearer if the rules of description are clearly defined in the internal evaluation manual, etc.)



This report is a compilation of the results of a study conducted by KPMG AZSA LLC on behalf of the Japan International Cooperation Agency. We strive to timely compile this report based on the information available at the time of study. However, the content of this report does not necessarily correspond to situations of certain individuals or organizations those are not included in this study. In addition, this report does not guarantee the accuracy or completeness of the information at the time of receiving the information or thereafter. This report has been submitted only to the Japan International Cooperation Agency. KPMG AZSA LLC assumes no direct or indirect responsibility for the use of this report by third parties who have viewed or obtained a copy of this report.