

# **Evaluation Capacity Development for Educational Development: Case Study of the Policy-level Evaluations of Japan's ODA in the Education Sector**

Yoko Ishida  
*Hiroshima University, Japan*

## **Abstract**

Education policies and systems have been experiencing increasing pressure to show greater accountability and effectiveness (OECD, 2007). In Japan, under the School Education Law and the Government Policy Evaluation Act, thousands of evaluations are conducted at various levels of education and educational administration every year. If the results are not used, there would be limited meaning in doing evaluation. In order to strengthen the use of evaluations, the quality of evaluations needs to be enhanced to meet stakeholders' needs. The study used meta-evaluation of Japan's educational cooperation policies to understand how evaluation results are used and what factors are hindering. Japan's new education cooperation policy states that Japan provides support in light of "Japan's strengths in its education system," which have still been vaguely understood. To create evidence to clarify "Japan's strengths in its education system" is expected role of evaluation. The conceptual framework for administering and using evaluations, as well as some thoughts about its potential uses and applications should be developed by providing research on evaluation with practical case studies to improve quality of evaluations and evaluation results.

## **Introduction**

In the past 10 years, official evaluation systems were introduced to various levels of educational development in Japan, where policy-makers, central and local government administrators, head teachers, teachers, students, and guardians are requested to do and participate in evaluation activities and/or to be put in a position to be evaluated.

The School Education Law, amended by MEXT (the Ministry of Education, Culture, Sports, Science and Technology-Japan) in 2007, requires elementary, lower secondary, and higher secondary schools to conduct an internal evaluation of their management and performance annually and share findings for improving mutual understanding and collaboration with guardians and community members (MEXT, 2016). Universities have been requested to complete certified evaluation and accreditation by National Institution for Academic Degrees and Quality Enhancement of Higher Education (NIAD-QE, 2016) under the national evaluation scheme introduced by MEXT in 2004, in addition to

conducting self-evaluations of management and performance and reporting findings to the public. At the policy level, in addition to their evaluations of ODA (official development assistance) policies and programs, MEXT and MoFA (Ministry of Foreign Affairs) have been required since 2003 to conduct evaluations of their administrations, following the Government Policy Evaluations Act adopted in 2001 (MIC, 2001).

Weiss (1998) defines evaluation as the systematic assessment of the operation and/or the outcomes of a program or policy compared to a set of explicit or implicit standards, as a means of contributing to the improvement of the program or policy. In the context of international cooperation, OECD DAC (Development Assistance Committee) (2002) defines evaluation as the systematic and objective assessment of an on-going or completed project, program or policy, its design, implementation and results, which aims to determine the relevance and fulfillment of objectives, development efficiency, effectiveness, impact and sustainability. It also states that evaluations should provide information that is credible and useful and enable the incorporation of lessons learned into the decision-making process of both recipients and donors (OECD 2002).

As in this definition of evaluation, evaluation aims not only to prepare a report about findings and recommendations but also to use evaluation results to improve programs or policies. The 2000s have seen a movement toward the use of evaluations increasingly for the purpose of accountability rather than for management or learning, with a focus on government-wide needs. For example, in Canada, the monitoring and evaluation system has encountered difficulty in finding the right balance in its use of tools to measure performance between their support management or learning, and their support of accountability to senior or central authorities (Lahey et al., 2013). In Japan, to fulfill the government requests, the evaluation systems at the government, university and school levels face the same type of difficulty as in Canada, giving top priority on completing evaluation duties and submitting documents rather than on using evaluations for organizational learning.

Since the beginning of the 2000s, with the government administration reform under the concept of NPM (new public management), result-based management has been promoted in Japan. Result-based management and results orientation in the international cooperation has shifted an increased emphasis onto policy-makers and program- and project-managers to measure outputs and outcomes from their interventions. By pointing to the growing focus on evaluation and assessment, OECD (Organization for Economic Co-operation and Development) (2013) states:

Results from evaluation and assessment are becoming critical to knowing whether the school system is delivering good performance and to providing feedback for further development. Evaluation and assessment are instrumental in defining strategies for improving practices within school systems with the ultimate goal of enhancing student outcomes. These developments are having a strong influence in the way in which policy makers monitor system, school, school leader, teacher and

student performance. (p.17)

There are guidelines and handbooks developed on how to design and conduct evaluations prepared by responsible agencies; however, these resources allocate few pages to explain how to disseminate and use the results of the evaluation. Most work in the area of capacity development in education focuses on the supply side (capacity to do evaluation), and little attention has been paid to the demand side (capacity to use evaluation) (Cousins, 2014). Every year, various efforts are made to collect data and prepare documents by government officers, head teachers and teachers, who are expected to use evaluations for developing their organizational learning capacity.

To enhance the use of evaluations, some international agencies and bi-lateral donor agencies have introduced a system called management response. For example, UNICEF (United Nations Children's Fund) (2012) states that a management response is required for every evaluation, where fully or partially conducted by UNICEF, following procedures that include disseminating evaluation reports and stimulating stakeholder interest and preparation of management response documents about all actions to be undertaken based on evaluation recommendations. These action plans and follow-up actions are checked by UNICEF's own tracking system.

In the area of Japan's ODA, MoFA conducts between 8 and 14 third-party evaluations per year (MoFA, 2016), and JICA (Japan International Cooperation Agency) has conducted 70-100 third-party ex-post evaluations (JICA, 2016). Supposing that it costs approximately 5 million yen to conduct a third-party evaluation by hiring consultants, around 500 million yen are invested every year for evaluations. It is not the intention of this study to adjudicate whether this investment is too high or too low. For realizing the optimum use of invested money, evaluation findings should be used not only for accountability purposes, but also to learn from experience. In order to encourage effective use of evaluations, it is necessary to rethink how evaluation outputs are created to meet the needs of evaluation users (policy-makers, central and local governments, head teachers, teachers, and citizens).

## **Study Objective**

The study aimed to understand how evaluation results are used by the stakeholders for improving planning and implementation process of educational development and to identify factors that hinder stakeholders from relying on and/or using evaluation results.

The study questions are i) how and by whom evaluation results and recommendations are created and used for organizational learning; and ii) what are hindering factors and possible measures to improve evaluation process and outputs.

## **Significance of the Study**

There have been thousands of evaluations carried out at the school, university and central and local educational administration levels every year for the past 10 years in Japan, which has cost billions of yen. If the results are not used, spending so much money seems meaningless.

In order to strengthen the use of evaluations, the quality of evaluations need be enhanced. It is important to check evaluations for problems such as bias, technical error, administrative difficulties, and misuse, which reduces reliability of evaluations. Such checks are useful both to improve ongoing evaluation activities and to assess the merits of completed evaluation efforts (Stufflebeam, 2011).

By conducting meta-evaluation of the completed policy-level evaluations in the education sector, the study checked their level of technical adequacy and utility, which would provide substantial data to improve the meta-evaluation methodology as well as to value data produced by evaluations, which leads to integration of evaluations into organizational culture.

## **Methodology**

A case study of MoFA's evaluations of ODA policies for educational development was performed using the meta-evaluation check criteria provided by Scriven (2015).

Scriven (1969) introduced the term "meta-evaluation" to refer to evaluation of evaluations. This approach appraises evaluations based on three standards: technical adequacy, utility and cost/effectiveness for improving on-going evaluation activities, and for assessing the merits of completed evaluation efforts (Stufflebeam, 2011). Under these three standards, Stufflebeam (2011) provides 11 criteria, including internal validity, external validity, reliability, objectivity, relevance, importance, scope, credibility, timeliness, pervasiveness, and cost/effectiveness. In his meta-evaluation checklist, Scriven (2015) enumerates eight items: validity; credibility; clarity; concision; timeliness; effectiveness; cost-feasibility and cost-effectiveness, which must be determined through meta-evaluation.

Six criteria from Scriven's checklist were selected as analysis viewpoints for this study in order to evaluate the technical adequacy and utility of evaluations: validity, credibility, clarity, concision, timeliness, and effectiveness. The cost-feasibility and cost-effectiveness criteria were not chosen for the study because these items do not directly relate to how evaluations are used and because cost information is not available from the evaluation reports.

Based on the findings of the meta-evaluation and the review of recommendations and response measures, possible hindering factors were examined.

## Research Target

In Japan, the evaluation systems of ODA policies, programs and projects have been strengthened by central government and implementation agencies. MoFA, as the policy-making agency with respect to ODA, is responsible for policy- and program-level evaluations (MoFA, 2013), while JICA, as an ODA implementation agency, conducts project evaluations.

MoFA's ODA evaluations have two objectives: i) learning in order to improve ODA management; and ii) accountability, including the provision of information to the public (MoFA 2013). In MoFA, after conducting evaluations, the ODA Evaluation Division of MoFA invites the International Cooperation Bureau as well as JICA and hosts an internal follow-up meeting in order to report evaluation results and develop response measures for the International Cooperation Bureau (MoFA, 2013). MoFA publishes an annual evaluation report for showing the third-party evaluation results and how they follow up on recommendations as response measures.

The study targeted the three evaluations of Japan's ODA policies and initiatives in the education sector conducted by MoFA, as follows: "Mid-term Evaluation on Japan's Contribution to the Achievement of the MDGs in the Area of Education (hereinafter, referred to as Mid-term Evaluation)" in JFY2004; "Evaluation of Japanese Educational Cooperation Policy: Basic Education for Growth Initiative (hereinafter, referred to as BEGIN Evaluation)" in JFY2007; and "Evaluation on Japan's Education Cooperation Policy 2011-2015 (hereinafter, referred to as JECP Evaluation)" in JFY2015.

These three evaluations were carried out to assess the progress and achievement of each of the target policies or initiatives and to understand Japan's contribution to the achievement of EFA (Education for All) and MDGs (Millennium Development Goals) in the area of education. The preceding evaluation results and recommendations are expected to be reflected in subsequent policy-making. The three evaluations are outlined and compared as shown in Table 1, including their targets, background, objectives, methods, results (rating) and so on.

**Table 1. Correspondence among the MoFA Three Evaluations**

Items	Mid-term Evaluation	BEGIN Evaluation	JECP Evaluation
1. Target of evaluation:	Japan's contribution to the MDGs in the area of education	Japan's educational cooperation initiative: <i>BEGIN</i>	<i>Japan's Education Cooperation Policy 2011-2015</i>
2. Goals (indicators) of target policies / initiatives:	Goals were not set	Goals were not set	-US\$3.5 billion input in the field of education -At least 7 million children (a cumulative total of around 25 million children) supported

3.	Sub-sectors of target policies / initiatives	Basic education	Basic education	Basic education, upper secondary education and higher education
4.	Evaluation implemented in:	JFY2004	JFY2007	JFY2015
5.	Background:	-The year before the international MDGs mid-term evaluation	-Midpoint of the MDGs period -The year before Japan became co-chair of EFA-FTI (Fast Track Initiative)	-MDGs and EFA target year -Final year of <i>Japan's Education Cooperation Policy</i>
6.	Type of evaluation:	Third-party evaluation	Third-party evaluation	Third-party evaluation
7.	Evaluation (consultant) team:	KRI International	International Development Center of Japan	International Development Center of Japan
8.	Objectives:	-To learn for further development of contribution -To fulfill accountability	-For better policy formulation and more effective and efficient implementation of ODA in basic education -For transparency and accountability	-To make vital recommendations based on findings and lessons from the evaluation -To fulfill accountability
9.	Methods:	1) Designing of framework 2) Literature and statistics review 3) Interview with persons concerned 4) Case study in Vietnam	1) Designing of framework 2) Literature and statistics review 3) Interview with persons concerned 4) Questionnaire to Japanese embassies 5) Case study in Kenya and Ethiopia	1) Designing of framework 2) Literature and statistics review 3) Interview with persons concerned 4) Questionnaire to Japanese embassies 5) Case study in Senegal 6) Meta-evaluation of JICA evaluation reports
10.	Rating by Evaluation Team:			
10-1	Relevance of policies:	-Japan's assistance approaches are relevant to the MDGs regarding education.	- <i>BEGIN</i> is relevant to the ODA Carter. - <i>BEGIN's</i> role and objectives are not clear	-The relevance of the policy is high.
10-2	Effectiveness of results:	-Japan's assistance for increasing access has been effective, but its assistance related to literacy and gender needs to be improved.	-Japan's assistance for quality improvement and ensuring access has been effectively done	-The overall effectiveness of the policy is high.

10-3 Appropriateness of process:	-Rating is not clearly provided.	- <i>BEGIN</i> is not recognized as a practical strategy, but as an idealistic goal.	-The policy was implemented in a moderately satisfactory manner.
10-4 Diplomatic viewpoints:	-Rating is not clearly provided.	- <i>BEGIN</i> has not been effective from the diplomatic viewpoint.	-The policy has diplomatic importance and diplomatic impact.

## Findings of Meta-evaluation

The reports of the three policy-level evaluations were reviewed with respect to the six (6) meta-evaluation check items of i) validity, ii) credibility, iii) clarity, iv) concision, v) timeliness, and vi) effectiveness selected from the checklist provided by Scriven (2015).

### *Validity*

Validity is the criteria of technical adequacy (Scriven, 2015). To examine validity, the following points are considered: 1) relevance of evaluation design; 2) appropriateness of evaluation methodology and implementation process including ethicality, legality and propriety of evaluation process; and 3) accuracy of evaluation results.

In the three evaluations, all evaluation activities were designed and implemented strictly following the ODA evaluation guidelines prepared by MoFA. The same evaluation design, methodology, and implementation processes were applied from the three evaluation viewpoints: a) relevance of the policy and/or initiative; b) effectiveness of the results; and c) appropriateness of the processes. Evaluation frameworks were prepared at the beginning of the evaluation, including indicators, evaluation questions and possible data sources for each of the three evaluation viewpoints, and then evaluation activities were carried out according to the frameworks.

Overall, the three evaluations were conducted as third-party evaluations following the evaluation guidelines provided by the client, MoFA. The evaluation reports were technically well structured and prepared with ethics, legal and propriety matters carefully considered.

As for the accuracy of evaluation results, because outcome goals were not set for any of the policies and initiatives evaluated, the judgements of the effectiveness done by the evaluation teams are subjective. Most of the analysis was done based on Japan's ODA input (amount of money) and project and program performance in the form of descriptive data, both of which were collected from ODA-concerned parties and persons. Clear reasons for the judgements with objective evidence of the outcome or impact were seldom given; therefore, none of the three evaluations report to what extent the goals were attained or how much Japan's ODA contributed to the achievement of the international goals of education development.

### ***Credibility***

This criterion relates to the utility of evaluation and concerns whether the audience trusts the evaluator and supposes him/her to be free of bias in his/her conduct of the evaluation (Stufflebeam, 2011). Audiences are often not in a position to assess the technical adequacy of a study; the next best thing they can do is to decide whether they have confidence in the group that conducted the study and in their judgement and objectivity. This factor is often correlated with the matter of independence (Stufflebeam, 2011).

The three evaluation reports were created professionally, and there is no appearance of an inadequate level of competence and no apparent conflicts of interest between the client (MoFA) and the evaluators observed, which are positive points in favor of their credibility. However, as mentioned in the sub-section on validity, the judgement results of the three evaluations were given without showing any objective evidence. These situations made the three evaluation results less persuasive as grounds for the recommendations produced by these evaluations.

### ***Clarity***

Clarity is another criterion of utility, which concerns dissemination and use of the evaluation findings as well as fulfilling of accountability. This is required for all stakeholders with a right to know; this may require doing more than one report, written or spoken; and it commonly required post-report interaction with audiences in order to check for misunderstanding (Scriven, 2015).

Following the MoFA guidelines of how to prepare ODA evaluation reports, the three evaluation reports were prepared carefully, considering comprehensibility for a general audience. In spite of the intentions of the evaluators, many jargons used in the international development issues; such as grant aid, gender consideration, donor coordination and so on, may hinder stakeholders' interest in reading and using evaluation reports. These jargons are necessary for the evaluators to create reports with technical adequacy; however, the jargons might make many stakeholders believe that international cooperation for educational development is something special and very different from what they experience at school in Japan.

Data of Japan's input, how much amount of ODA was invested, was provided; however, there are limited quantitative outcome and impact data shown as evidence of Japan's contribution. The reasons of the rating are given based on qualitative information from the individual projects, which makes the evaluation results less convincing.

### ***Concision***

Concision is also a criterion of utility, including what is required for validity and utility while avoiding unnecessary length (Scriven, 2015).

When talking about the report length, the three evaluation reports were created with an appropriate length of approximately 100 pages, following the MoFA guidelines of how



to prepare ODA evaluation reports.

The three evaluation reports have a preference for circumlocutious (or circumlocutory) expressions for describing the evaluation results and the reasons, which might have been caused by the limited availability of outcome and impact data. Roundabout expression with many jargons may negatively affect stakeholders' interest in and reliance on the evaluation results.

### ***Timeliness***

Timeliness is perhaps the most critical of the utility criteria (Stufflebeam, 2011). An evaluation that comes in after the decision deadline is often of zero value, not just of less value (Scriven, 2015).

Mid-term Evaluation was conducted to prepare necessary data for the international MDGs mid-term evaluation, which was planned in the following year. BEGIN Evaluation was done to evaluate Japan's education cooperation initiative and to understand the progress of Japan's contribution

To the attainment of the international development goals at the midcourse of EFA (Education for All) and MDGs (Millennium Development Goals) regarding educational development. Considering the backgrounds of these two evaluations and their recommendations used in the succeeding policy-making, the two evaluations were conducted in a timely manner.

JECP Evaluation was performed in JFY2015, and the reports were produced in March 2016 as originally scheduled; however, the new education cooperation policy, "Learning Strategy for Peace and Growth – Achieving Quality Education through Mutual Learning", was announced at the UN (United Nations) Sustainable Development Summit in September 2015. It is clear that the JECP evaluation was not designed to provide lessons or recommendations for the succeeding cooperation policy, and it is not clear how MoFA plans to use the results of the untimely evaluation.

### ***Effectiveness***

Scriven (2015) states that the criterion of effectiveness is to be judged until the close of the follow-up window and that it includes utilization-maximizing design and execution/assistance as well as the outcomes of the evaluation.

When checking how the evaluation results and recommendations were reflected in the policy-making, among the three evaluations, the author found the following cases in which recommendations were reflected in the succeeding policy-making and measures were taken for improvement during the policy implementation:

- Mid-term Evaluation and BEGIN Evaluation recommended that Japan should make visible and useful contributions to FTI and, reflecting the recommendation, the succeeding policy of Japan's Education Cooperation Policy 2011-2015 included "strengthening of Japan's support for FTI" as one of its focus areas. JECP Evaluation positively evaluated Japan's contribution to FTI.

- Mid-term Evaluation and BEGIN Evaluation recommended that an educational development assistance policy (initiative) should include verifiable targets and goals in order to assure timely and effective realization of the policies. The target goals of Japan's input (amount of money) and the number of children covered by Japan's support were set up for Japan's Education Cooperation Policy 2011-2015. JECPEvaluation assessed this setting-up of the goals positively, even they were input goals, and pointed out the necessity of setting up outcome and impact goals as well as of strengthening a monitoring system for policy implementation.

There are also less effective cases, where recommendations are not properly reflected in the succeeding policy-making or no improvements have been observed.

- Mid-term Evaluation and BEGIN Evaluation recommended that MoFA should clearly define official roles of an educational development assistance policy (initiative) as principal education cooperation policy in Japan's ODA. Follow-up activities to the recommendations were shown as response measures in MoFA's annual evaluation report. However, according to the questionnaire survey of JECPEvaluation, only 20% of the Japanese embassies regarded Japan's Education Cooperation Policy 2011-2015 as Japan's principal education policy and took any action to promote the policy in their responsible countries.
- Mid-term Evaluation and BEGIN Evaluation recommended that a monitoring system should be strengthened and monitoring activities should be implemented following a properly designed monitoring and evaluation plan. Strengthening of a monitoring system was included as one of the key strategies in Japan's Education Cooperation Policy 2011-2015. However, JECPEvaluation found that the monitoring system had not been functional or properly implemented during the policy implementation period.
- Mid-term Evaluation and BEGIN Evaluation recommended that a cooperation policy should have a practical annual plan of action and a functional section/unit of MoFA assigned for policy implementation and monitoring; however, no action plan or implementation structure was developed for Japan's Education Cooperation Policy 2011-2015, and none has yet been developed for the succeeding education cooperation policy.

## **Conclusion**

Factors and conditions supporting or impeding organizational uses of evaluations include at least the following: timeliness, constructive nature of feedback, information needs of primary users, credibility of findings, accessibility to primary users, communicability, involvement of non-evaluator stakeholders, and relevance to decision priorities (Cousins, 2014).

Based on the meta-evaluation results of this study, among the factors and conditions

mentioned above, *information needs of primary users* and *involvement of non-evaluator stakeholders* need be improved. When recommendations and response measures need to be handled by MoFA sections, which are responsible for preparing and/or evaluating policies, then it is easier for those sections to reflect and practice them. However, when stakeholders who do not belong to MoFA planning and evaluation sections, such as the MoFA ODA implementation section, Japanese embassies etc., are involved, their response measures are apt to be postponed.

It has been shown by some research results that the successful use of evaluation data in organizations fosters their valuing by members as a powerful force for organizational and program change (Cousins et al., 2005). Organizational decision makers need to experience the benefits of evaluation firsthand before they willingly embrace it as leverage for change (Cousins et al., 2014).

As shown in the findings of this study, evaluation is an intervention and a process, one that requires the facilitation of interactions at multiple steps along the way (Catsambas, 2016). The evaluator engages in facilitation throughout the course of an evaluation to promote the participation of non-evaluator stakeholders, including not only central and local government officers at the central level but also head teachers, teachers and guardians at the school level, for effective use of evaluations.

The author concludes that, in order to improve the quality of evaluations and deliver evaluation results that meet the needs of evaluation users, the conceptual framework to perform evaluations, as well as their potential uses and applications, should be developed further by providing research on evaluations with practical case studies involving meta-evaluation and organizational evaluation capacity development.

## References

- Bandstein, S., & Hedblom, E. (2008). *IFAD's Management Response System: The Agreement at Completion Point Process*. SADEV Report 2008:4. Stockholm: Swedish Agency for Development Evaluation (SADEV).
- Bandstein, S., & Hedblom, E. (2008). *Reaping the Fruits of Evaluation? An evaluation of management response systems within aid organizations*. SADEV Report 2008:7. Stockholm: SADEV.
- Catsambas, T. T. (2016). Facilitating Evaluation to Lead Meaningful Change. In R. S. Fierro, A. Schwartz, & D. H. Smart (Eds.). *Evaluation and Facilitations. New Directions for Evaluation*, 149, 19-29.
- Consins, J. B., Bourgeois, I., & Associates (2014). Cross-Case Analysis and Implications for Research, Theory, and Practice. In J. B. Cousins & I. Bourgeois (Eds.). *Organizational capacity to do and use evaluation, New Directions for Evaluation*, 141, 101-119.
- Consins, J. B., Bourgeois, I., & Associates (2014). Multiple Case Study Methods and Findings. In J. B. Cousins & I. Bourgeois (Eds.). *Organizational capacity to do and use evaluation, New Directions for Evaluation*, 141, 25-99.

- Consins, J. B., Goh, S. C., Elliott, C. J., & Bourgeois, I. (2014). Framing the capacity to do and use evaluation. In J. B. Cousins & I. Bourgeois (Eds.). *Organizational capacity to do and use evaluation, New Directions for Evaluation*, 141, 7-23.
- Cousins, J. B., Goh, S., & Clark, S. (2005). Data use leads to data valuing: Evaluative inquiry for school decision making. *Leadership and Policy in Schools*, 4, 155-176.
- Cousins, J. B., Goh, S., Clark, S., & Lee, L. (2004). Integrating evaluative inquiry into the organizational culture: A review and synthesis of the knowledge base. *Canadian Journal of Program Evaluation*, 19(2), 99-141.
- Fullan, M. (2016). *The New Meaning of Educational Change, Fifth Edition*. New York, U.S.A.: Teachers College Press.
- Hanberger A., & Gisselberg K. (2006). Sida's Management Response System. *Sida Studies in Evaluation 06/01*. Stockholm: Swedish International Development Cooperation Agency (SIDA).
- Hatry, H. P., (2013). Sorting the relationship among performance measurement, program evaluation, and performance management. In S. B. Nielsen & D. E. K. Hunter (Eds.). *Performance management and evaluation, New Directions for Evaluation*, 137, 19-32.
- Iwasaki, K. (2010). Evidence-based policies in education: new directions and issues. *Japanese Journal of Evaluation Studies*, 10(1), 17-29.
- JICA (2016). *Evaluation, Operations Evaluation in JICA*. Retrieved from [http://www.jica.go.jp/english/our\\_work/evaluation/](http://www.jica.go.jp/english/our_work/evaluation/) (Accessed on August 4, 2016).
- Lahey, R., & Nielsen, S. B., (2013). Rethinking the Relationship Among Monitoring, Evaluation, and Results-Based Management: Observations from Canada. In S. B. Nielsen & D. E. K. Hunter (Eds.). *Performance management and evaluation, New Directions for Evaluation*, 137, 45-56.
- MEXT (2016). *School Evaluation, School Education Law and School Evaluation*. Retrieved from [http://www.mext.go.jp/a\\_menu/shotou/gakko-hyoka/08021216.htm](http://www.mext.go.jp/a_menu/shotou/gakko-hyoka/08021216.htm) (Accessed on August 5, 2016).
- MIC (Ministry of Internal Affairs and Communications) (2002). *Government Policy Evaluations Act (Act No.86 of 2001)*. Tokyo: MIC.
- Minamoto, Y. (2013). Evaluating Evaluations (Meta-evaluation), *In MIC The Study on the Policy Evaluation Checking Systems of the Foreign Countries*. Tokyo: MIC.
- MoFA (2004). *Mid-term Evaluation on Japan's Contribution to the Achievement of the MDGs in the Area of Education*. Tokyo: MoFA.
- MoFA (2006). *Annual Evaluation Report on Japan's Economic Cooperation 2006*. Tokyo: MoFA.
- MoFA (2007). *Evaluation of Japanese Educational Cooperation Policy "Basic Education for Growth Initiative."* Tokyo: MoFA.
- MoFA (2009). *Annual Evaluation Report on Japan's Economic Cooperation 2009*. Tokyo: MoFA.
- MoFA (2013). *ODA Evaluation Guidelines 8<sup>th</sup> Edition*. Tokyo: MoFA.
- MoFA (2015). *Learning Strategy for Peace and Growth – Achieving Quality Education*

- through Mutual Learning* -. Tokyo: MoFA.
- MoFA (2016). *Evaluation on Japan's Education Cooperation Policy 2011-2015*. Tokyo: MoFA.
- MoFA (2016). *Japan's ODA, ODA evaluation*. Retrieved from <http://www.mofa.go.jp/policy/oda/evaluation/index.html> (Accessed on August 3, 2016).
- NIAD-QE (2016). *University Evaluation, NIAD-QE's Evaluation and Accreditation*. Retrieved from <http://www.niad.ac.jp/english/unive/> (Accessed on August 5, 2016).
- OECD (2007). *OECD Evidence in Education: Linking Research and Policy*, Paris; OECD.
- OECD (2013). *OECD Reviews of Evaluation and Assessment in Education, Synergies for Better Learning, An International Perspective on Evaluation and Assessment*, Paris; OECD.
- OECD-DAC (Organization for Economic Co-operation and Development-Development Assistance Committee) (2002). *Glossary of Key Terms in Evaluation and Results Based Management*, Paris; OECD.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation, A Systematic Approach, Seventh Edition*. Thousand Oaks, California, U.S.A.: SAGE Publications, Inc.
- Sasaki, R. (2010). An In-Depth International Comparison of Major Donor Agencies: How Do They Systematically Conduct Country Program Evaluation? *Journal of MultiDisciplinary Evaluation*, 8(18), 29-46.
- Scriven, M. S. (1969). An Introduction to Meta-Evaluation. *Educational Product Report*, 2(5), 36-38.
- Scriven, M. S. (2015). *The Meta-Evaluation Checklist, New Version 11.13.2015*. Retrieved from <http://michaelscriven.info/papersandpublications.html> (Accessed on August 4, 2016).
- Stufflebeam, D. L. (2001). Evaluation Models. *New Directions for Evaluation*, 84, 7-98.
- Stufflebeam, D. L. (2011). Meta-Evaluation. *Journal of MultiDisciplinary Evaluation*, 7(15), 9-158.
- Sturges, K. M. (2015). Complicity Revisited: Balancing Stakeholder Input and Roles in Evaluation Use. *American Journal of Evaluation 2015*, 36 (4), 461-469.
- UNICEF Evaluation Office (2012). *Guidance for Management Response to Evaluations: Enhancing Critical engagement for the Strategic Use of Evaluations*, New York: UNICEF.
- Weiss, C. H. (1998). *Evaluation, Methods for Studying Programs and Policies, Second Edition*. Upper Saddle River, New Jersey, U.S.A.: Prentice-Hall, Inc.
- Yutaka, K. (2011). Roles and Issues of the Agency Working with Evidence – the example of What Works Clearinghouse in the United States. *Bulletin of the National Institute for Educational Research*, 140, 71-93.