【研究論文】

A Consideration of Impact Evaluation Methodology at the Sector Level – A Case Study Using Donors' Assistance Projects in Primary and Secondary Education in Indonesia Since 1990

Satoshi Morita, Tetsuya Araki, Yasuyuki Sagara
Graduate School of Agricultural and Life Sciences, the University of Tokyo
morita.satoshi@alumni.clarku.edu

Abstract

This paper proposes a framework for impact evaluation of donors' assistance projects at the sub-sector level, using the case of PSE in Indonesia. It presents: 1) methodological procedures in four stages, using a program theory model and an interrupted time series model in impact evaluation; 2) types of necessary data for analysis; and 3) analysis results based on collected data, focusing on higher outcome levels in the target sub-sector. It also identifies threats to the validity and adequacy of evaluation designs and results, and proposes some solutions for them. Finally, it reconsiders key issues observed in this study, such as: 1) the issue of contribution and attribution; and 2) donors' roles and responsibilities.

Keywords

impact evaluation, outcome, program theory model,
interrupted time series model, contribution and attribution

1. Introduction

Recently, some limitations of project-type assistance approaches have been pointed out, and many discussions on a shift to program approaches have been held. Since the mid-nineties, a sector-wide approach has been applied in some developing countries in which various donors provide their support based on development plans in certain sectors in coordination with the recipient country's government. Another example is the "common basket" approach aimed at standardization and rationalization of different donors' assistance frameworks under a certain common goal by pooling all available funds (IDJ2004). In 1999, the World Bank (WB) proposed the Comprehensive Development Framework (CDF), an assistance approach for addressing many development issues for the target period of 10 to 15 years with various donors' participation and involvement. While poverty reduction has been drawing more attention, the International Monetary Fund and the WB, aiming at embodying the concept of CDF, request heavily indebted countries to develop a Poverty Reduction Strategic Paper to promote effective support in cooperation with various donors (MOFA2003).

This shift to program approaches and the tendency towards assistance coordination among donors highly relate to the recent trends of result orientation. That is, based on recent reforms in the management systems at donor agencies, they put emphasis on performance measurement using indicators and numerical values, focusing on "output," including products, capital and services in assistance programs and projects, and "outcome," including changes in target groups in the recipient country. In this system, the attainment of shared goals at the sector level, or the attainment of outcomes is emphasized. Some donors initiated efforts to introduce a results-based management system at an earlier stage and established their strategic plans and country-based assistance plans relating to internationally shared goals such as Millennium Development Goals.

At these agencies there have been some cases of evaluation in which attempts were made to observe to what degree situation changes at the country level were attributed to their own assistance projects in a recipient country. In the meantime, project evaluation remains the main trend at donor agencies, since most of them implement individual projects-based assistance. So far, there are very few cases of evaluation studies focusing on impact, or long term effects at higher outcome levels, produced by all relevant assistance projects conducted by various donors in a target sector in a recipient country. Therefore, no systematic method has been established.

There exist some technical difficulties in observing impact of assistance projects at the sector level, such as: 1) it may require substantial costs and time to gather information since a large amount of information is dealt with (resources issue); 2) it is not easy to objectively determine the attribution of the appearance of impact to projects since many factors are involved (methodological issue); and 3) evaluators' specialties regarding the target sector or country may be required (skills issue). These issues may be hindering factors in conducting evaluation case studies. However, the authors assume it is possible to overcome these difficulties depending on the method to be applied. Based on this assumption, this paper reconsiders impact evaluation methodology from a practical viewpoint, using a case study in which one can see what effects are produced at the higher outcome level in a set time period by various donors' projects conducted in a target sector in a recipient country.

2. Framework for Impact Evaluation Methodology

This paper first proposes unique viewpoints and a basic framework for research and analysis methodology, referring to the discussions of Bamberger *et al.* (2004) regarding project impact evaluation designs. Table 1 shows the basic framework, composed of four stages.

3. Making Evaluation Plans

3.1 Selecting Evaluation Target, and Setting the Evaluation Purpose and Question

The first stage is to make evaluation plans. In this study, primary and secondary education (PSE) in Indonesia is selected as a target sub-sector. The evaluation purpose is "the examination of effects produced by donors' assistance projects in PSE implemented under the centralized administrative systems since 1990," and the evaluation question is as follows: "What effects have these individual projects produced, as a whole, at the outcome level in the entire country[1]?" Table 2 shows a list of 19 main projects implemented or funded by various donors in PSE since 1990[2]. It shows target regions, project period, input (grant or loan), target education levels and activities for each project. There are 27 categories in the target region, three in the education level (PR, JS and SS) and 51 in the activity (a to Y), as shown in Table 3 (see later).

Table 1 Basic Framework for Research and Analysis Methodology

Stages	Contents
1 Making Evaluation Plans	*1 Selecting Evaluation Target, and Setting the Evaluation Purpose and Question* 2 Reconfirming Policies Based on Program Theory 3 Making Tentative Evaluation Designs
2 Identifying Constraints in Research and Determining Evaluation Designs	1 Identifying Budget, Time and Data Constraints 2 Determining Evaluation Designs - *Setting Criteria and Identifying Core Outputs and Activities* - Considering Ways to Reduce the Number of Survey Targets *and Identifying Data Collection Sites* - Considering Methods of Data Collection and Minimizing Cost and Time
3 *Implementing Evaluation Surveys and Analyzing Results*	*1 Implementing Evaluation Surveys* - *Reconfirmation of Program Theory Model by Policy-makers* - *Data Types and Collection Results* *2 Analyzing Survey Results* *3 Summary of Evaluation Results* *4 Consideration of Evaluation Results*
4 Identifying Threats to the Validity and Adequacy of Evaluation, *and Proposing Solutions*	1 Validity and Adequacy of Evaluation Designs 2 Validity and Adequacy of Evaluation Results

Source: The authors, based on Bamberger *et al.* (2004)
Note: Items in italic are the authors' original ideas.

3.2 Reconfirming Policies Based on Program Theory

Causal relationships between relevant factors in the government policies in the target sector are to be ascertained based on program theory. It becomes necessary to provide interpretation of the policies, based on literature reviews, as a program theory model. In this study, the Program Theory Matrix (PTM) is applied as an analysis tool[3] and the policies in PSE are presented in Table 3, based on the authors' interpretation regarding the contents of the country's national development plans including former REPELITAs and current PROPENAS, strategic plans of the Ministry of National Education (MoNE) in PSE, and other relevant materials. In the case that the target evaluation period is assumed for a mid-/long term, different models may be defined with separate time spans each time any change is observed in the factors or the causal relationships. In this study, however, the authors recognized that there had been no change in the basic structure of the causal relationships[4], even though some important policy goals were newly set and exogenous factors appeared in the target period in line with the identical causal relationships. Thus, only one program theory model is used in this study.

Table 3 shows: 1) one end outcome (EO) － i.e. improvement in the quality of PSE[5] (or improvement in the students' academic skills) in Indonesia; 2) four intermediate outcomes (IOs), including improvement in the quality of teachers, improvement in administrative functions, realization of universal education for students, and improvement in the school environment of students; and 3) one to four outputs for each IO. For instance, IO1 consists of three outputs, such as: 1) strengthening and improving the functions of PPPGs/BPGs/PKGs; and 2) those of LPTKs; and 3) stabilization of teachers' employment and dispatch systems. That is, the achievement of the three outputs will lead to IO1 and, in turn, the achievement of the four IOs will eventually lead to the EO. Also, Table 3 shows 51 activities (a to Y) and the positioning of the 19 projects containing one or several of the activities as their components[6]. Some issues in other sectors are also observed, such as governance improvement in IO2, poverty reduction in IO3, and infrastructure development in IO4.

Table 2 A List of Donors' Projects in Primary and Secondary Education in Indonesia since 1990

No.	Donors	Projects	Target Regions	Periods	Inputs (mil. US Dollars)	Target Levels, Activities
1	WB	Second Secondary Education and Management*1	1-27	1990.2 -1997.12	154.2 69.2	JS, SS/ a, c, e, u, v, w, z, C
2	WB	Primary Education Quality Improvement	1, 3, 12, 14, 16, 21	1992.6 -1999.3	28.3 16.9	PR/ c, q, u, w, z, C, D, G, N, V
3	WB	Primary School Teacher Development	1-27	1992.9 -1999.6	36.6*2 0	PR/ f, g, i, k, l, m, n, q, s
4	WB	Book and Reading Development	1-27	1995.8 -2001.12	51.6 + 6.6*2 216.6	PR, JS/ c, w, D, L
5	WB	Secondary School Teacher Development	1-27	1996.5 -2001.1	60.4 + 1.1*2 26.7	JS, SS/ f, g, h, i, k, l, m, n, u, D, F, W
6	WB	Junior Secondary Education	13, 16	1996.10 -2004.6	87.3 + 23*2 55.4	JS/ c, e, q, r, u, v, w, z, B, C, D, G, I, K, N, O, P
			11, 12, 17, 18	1996.10 -2004.6	89 + 5.2*2 46.5	
			1, 3, 5, 6, 8	1997.1 -2004.6	92.5 + 24*2 67.5	
7	WB	Basic Education*1	10	1998.7 -2004.12	98.8 25.5	PR, JS/ c, e, h, m, q, s, v, w, x, z, B, D, G, I, N, O, S
			23, 25, 26	1999.5 -2006.4	63.8 6.1	
			2, 4, 7	1999.5 -2006.4	74.6 8.3	
8	ADB	Junior Secondary Education	1-27	1993.02 -1998.12	98.1 + 0.6*2 70.1	JS/ c, h, s, u, v, w, z, D, N, O, V
9	ADB	Senior Secondary Education	1-27	1995.10 -2000.10	72.6 + 0.6*2 43.7	SS/ a, b, c, d, e, u, v, w, z, A, B, D, I, N, O, V
10	ADB	Private Junior Secondary Education	8, 10, 13, 19, 23	1995.10 -2002.8	33.6 + 1.0*2 19.5	JS/ b, c, p, q, z, A, B, C, D, G, H, I, N
11	ADB	Basic Education	8, 10, 11, 13, 15, 19	1996.9 -2002.8	57.5 + 0.6*2 47.6	PR, JS/ a, c, h, s, w, z, A, B, D, E, F, H, I, N, O, P, V
12	ADB	Development of Madrasah Aliyahs (Islamic High Schools) *1	1-27	1997.3 -2003.4	85 + 0.6 63.4	SS/ c, h, s, w, x, z, A, B, D, H, I, K, L, N, O, S, V
13	ADB	Second Junior Secondary Education Project *1	19, 20, 21, 22, 24	1998.5 -2003.8	153.8 49.7	JS/ c, h, q, r, w, z, B, D, I, K, M, N, O, P, S, Y
14	JBIC	Junior Secondary School Building Construction	4, 6, 8, 10, 11, 13, 15, 16, 17, 19, 23, 24	1995.12 -2000.12	120.7*3 5.5*3	JS/ N, O
15	MoFA/ JICA	Science and Mathematics Teaching for Primary and Secondary Education	10, 11, 13	1998.10 -2001.9	26.67 unknown	JS, SS/ f
16	JICA	Science and Mathematics Teaching for Primary and Secondary Education	10, 11, 13	1998.10 -2003.9	71.1*3 0.5*3	JS, SS/ f, g, h, i, l, m, n, E, U, V, X
17	JICA*4	Dispatch of Experts, JOCVs, and Senior Volunteers	–	–	–	PR, JS, SS/c, t, y
18	GTZ, KfW	Science Education Quality Improvement	9, 11, 13, 15, 19, 23, 27	1994.4 -2005.12	23.4*2*3 5.8*3	PR/ g, h, i, w, B, D, V
19	UNICEF, UNESCO, NZAid	Participatory Primary Education Improvement	11, 13, 16, 23	1996.6 -2002.12	4*2 unknown	PR/ K

Target Regions: 1. Aceh; 2. N. Sumatra; 3. W. Sumatra; 4. Riau; 5. Jambi; 6. S. Sumatra; 7. Bengkulu; 8. Lampung; 9. Jakarta; 10. W. Java; 11. Central Java; 12. Yogyakarta; 13. E. Java; 14. Bali; 15. W. Nusa Tenggara; 16. E. Nusa Tenggara; 17. W. Kalimantan; 18. Central Kalimantan; 19. S. Kalimantan; 20. E. Kalimantan; 21. N. Sulawesi; 22. Central Sulawesi; 23. S. Sulawesi; 24. Southeast Sulawesi; 25. Maluku; 26. N. Maluku; 27. Irian Jaya

Target Levels: PR(Primary); JS (Junior Secondary); SS (Senior Secondary)

Activities: See Table 3

Source: The authors, based on JICA(2001) and other relevant documents

Note: 1) Target regions, periods, inputs and activities of these projects (*1), as well as the numbers (*2) are shown as originally planned; 2) regarding inputs, numbers in the upper rows for loan projects (no. 1-14) are the amount of loans, while those for technical cooperation projects (no. 16-19) and grant aid project (no. 15) are the amount of grants. Numbers in the lower rows are the amount provided by the Indonesian government. Numbers (+) in the upper rows for the loan projects are the amount provided for relevant components in technical cooperation; 3) the amounts (*3) are calculated using approximate currency rates; and 4) JICA also provided several types of cooperation in different frameworks (*4) although data is not shown because people were dispatched to different regions at different times, and because the total budgets are estimated relatively low compared to the other projects due to the small number of people dispatched.

Table 3 Program Theory Model in Primary and Secondary Education in Indonesia (1990-2000)

End Outcome	Intermediate Outcomes	Outputs	Beneficiaries	Activities	Donors' Projects
improve the quality of primary and secondary education in Indonesia (improve students' academic skills)	1. improve the quality of teachers	1. strengthen and improve the functions of PPPGs, BPGs and PKGs	those concerned	a. establishment/improvement of facilities and equipment	1, 9, 11
			in-service teachers	b. improvement/development of educational programs	9, 10
				c. provision of re-training courses	1, 2, 4, 6, 7, 8, 9, 10, 11, 12, 13, 17
			trainers for teachers, etc.	d. strengthening program management	9
				e. re-training courses for trainers, lab. technicians, staff, etc.	1, 2, 6, 7, 9
		2. strengthen and improve the functions of LPTKs (former IKIPs, FKIPs and STKIPs) *LPTK stands for a teacher education institution.	overall	f. establishment/improvement of facilities and equipment	3, 5, 15, 16
			in-service teachers	g. improvement/development of re-training courses	3, 5, 16, 18
				h. provision of re-training courses	5, 7, 8, 11, 12, 13, 16, 18
			undergraduates (candidates for teachers)	i. improvement/development of educational programs	3, 5, 16, 18
				j. provision of education and training to undergraduates	-
				k. provision of relevant services to undergraduates	3, 5
			professors at LPTKs	l. strengthening research departments/activities	3, 5, 16
				m. provision of training courses	3, 5, 7, 16
			staff at LPTKs	n. strengthening program management	3, 5, 16
		3. stabilize teachers employment systems	in-service teachers	o. establishment of rewarding systems for quality teachers	-
				p. provision of sufficient salaries	10
				q. improvement of employment/dispatch systems	2, 3, 6, 10, 13
				r. establishment/improvement of residential facilities	6, 13
	2. improve administrative functions at DGPSE (Governance)	1. build personnel's capacities in policy formulation	administration (central/regional)	s. providing opportunities for training	3, 7, 8, 11, 12
				t. advisory provision	17
		2. build personnel's capacities in the development of curriculum, exam. systems and edu. materials	administration (central/regional)	u. strengthening functions at Curriculum Development Center	1, 2, 5, 6, 9
				v. strengthening functions at Examination Development Center	1, 6, 7, 8, 9
				w. strengthening functions at Book Center	1, 2, 4, 6, 7, 8, 9, 11, 12, 13, 18
		3. strengthen administrative systems	administration (central)	x. strengthening linkages between relevant departments	7, 12
				y. advisory provision	17
			(central/regional)	z. improvement of administrative management system	1, 2, 6, 7, 8, 9, 10, 11, 12, 13
			(regional)	A. development of school accredition systems	9, 10, 11, 12
			schools	B. strengthening school management	6, 7, 9, 10, 11, 12, 13, 18
		4. strengthen networks between central and regional administration offices and schools	administration (central/regional), schools	C. improvement of the information infrastructure	1, 2, 6, 10
				D. improvement of edu. materials distribution systems	2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 18
				E. holding seminars, workshops, conferences, etc.	11, 16
				F. dispatch of instructors to schools from LPTKs	5, 11
	3. realize universal education for students (poverty reduction)	1. guarantee provision of opportunities for education	schools	G. provision of subsidies (block grants, etc.)	2, 6, 7, 10
				H. provision of support for associations running schools	10, 11, 12
			parents, guardians, students	I. provision of scholarships to students	6, 7, 9, 10, 11, 12, 13
				J. reduction/exemption of educational costs	-
				K. awareness-raising campaigns for enrollment in compulsory edu.	6, 12, 13, 10
			students	L. awareness-raising campaigns for reading and learning	4, 12
				M. education programs outside school for drop-outs	13
	4. improve the school environment of students	1. improve education infrastructure	students	N. establishment/improvement of school facilities and equipment	2, 6, 7, 8, 9, 10, 11, 12, 13, 14
				O. establishment/improvement of school buildings/classrooms	5, 7, 6, 9, 11, 12, 13, 14
				P. establishment/improvement of dormitories	6, 11, 13
		2. improve students' commute conditions	students	Q. provision of support for commute fees	-
				R. improvement of the commute environment	-
		3. strengthen functions of school committees	parents, guardians	S. provision of support for PTA management	7, 12, 13
				T. reduction/exemption of parental associations fees	-
				U. development of school/classroom evaluation systems	16
	others	others	administration, schools	V. development of model schools (pilot activity)	2, 8, 9, 11, 12, 16, 18
			those involved in former IKIPs	W. provision of support for the transformation of IKIPs to universities	5
				X. public relations among former IKIPs	16
			in-service teachers	Y. publishing newsletter (MoNE's Provincial offices)	13

Source: The authors

Note: 1) PPPG stands for the Center for Teacher Training Development in different subjects. BPG is the Center for Teacher Training and PKG is the Peer In-service Teacher Training Program. There were 12 PPPGs nation-wide, 26 BPGs (one in each province) and numerous PKGs at the district level; 2) IKIP is the Teacher Training Institution with Education and other Faculty Departments, FKIP is the Faculty of Education of a university, and STKIP is the Teacher Training Institution with Education Faculty Department only. There used to be 10 IKIPs, 19 FKIPs and 2 STKIPs before 1999 when all IKIPs were transformed to universities.

3.3 Making Tentative Evaluation Designs

General evaluation designs are to be tentatively made, based on the program theory model. In this study, through utilizing the Evaluation Grid, research questions, necessary data items (or indicators), and expected information sources in the field are identified regarding the EO and the IO1-4, as shown in Table 4. Although necessary data items cover an extensive range, not all data items will be dealt with, since not all research questions in the proposed evaluation designs will be adopted in this study. It is important to first grasp the scope of potential field survey items. The survey targets will be addressed in more detail later.

4. Identifying Constraints in Research and Determining Evaluation Designs

4.1 Identifying Budget, Time and Data Constraints

The second stage is to identify in advance budget, time and data constraints in implementing evaluation research (Bamberger *et al.*2004). This research study was initiated for academic purposes, rather than by any request from a certain donor or recipient country. Therefore, the budget for this study was equivalent to the amount for two 10-day field surveys by an author in Indonesia. As research targets, this study deals with a number of projects since 1990, in which baseline surveys were not always conducted.

Table 4 Tentative Evaluation Designs

Research Questions: Outcomes	Research Questions: Outputs	Necessary Data Items (or Indicators)	Expected Info. Sources (in field surveys)
EO: Is the quality of primary and secondary education improved in Indonesia? (Are students' academic skills improved?)		*A.* international comparison of students' academic skills, results of national standard examination	MoNE
		B. assessment of primary and secondary education by experts	professors, experts
		C. students' habits and attitudes toward learning	schools, students
		D. the number of studetns per teacher	DGPSE
IO1: Is the quality of teachers improved?		*E.* the number of teachers meeting set quaifications	DGPSE
	Output1	*F.* the number and contents of re-training courses targeting in-service teachers, and the number of participants at PPPGs	DGPSE, target institutions, donors
		G. the number and contents of re-training courses at BPGs/PKGs	
		H. assessment of educational programs by teachers	participants (teachers)
	Output2	*I.* the number and contents of re-training courses targeting in-service teachers, and the number of participants at LPTKs	target LPTKs, donors
		J. assessment of new curriculum and syllabi at LPTKs	target LPTKs, donors
		K. the number and frequency of users of the facilities	target LPTKs, donors
		L. assessment of educational programs at LPTKs by students	participants (students)
		M. the number and contents of re-training courses targeting the faculty staff, and the number of participants at LPTKs	target LPTKs, donors
		N. assessment of the faculty staff by students	students at target LPTKs
	Output3	*O.* the number of teachers, employment statictics, salary systems	MoNE
IO2: Are administrative functions improved at DGPSE?		*P.* budgets/expenditures in primary and secondary education	DGPSE
		Q. efficiency/effectiveness of entire administration at DGPSE	DGPSE
	1, 2	*R.* assessment of personnel's skills and performance	DGPSE
	3	*S.* efficiency in administration, the number and contents of training courses and the number of participants	DGPSE, education dept. of regional governments
	4	*T.* effects of infrastructure development, the number of seminars/ workshops/conferences held and edu. materials distributed	DGPSE, regional governments, schools
IO3: Is universal education for students realized?		*U.* drop-out rates at each level	DGPSE
	1	*V.* enrollment rates, the number of scholarships/block grants provided	DGPSE, CSB
IO4: Is the school environment of students improved?		*W.* the average number of students per classroom	DGPSE
	1	*X.* the percentage of schools, classrooms, libraries that are usable	DGPSE
	2	*Y.* types of support for students' commute, average commute time and fees	schools, students
	3	*Z.* the number and contents of meetings and the number of participants, the degree of participation in school management	regional governments, schools, parents

Source: The authors

Contacting all the beneficiaries of the projects was not possible. Therefore, there was a limitation in data collection.

4.2 Determining Evaluation Designs

Evaluation designs are determined considering realistic feasibility of field surveys based on these constraints.

4.2.1 Setting Criteria and Identifying Core Outputs and Activities

The degree of realization of the EO and IOs, as well as the degree of their attribution to actual projects or activities are to be identified. Due to the various constraints, only one to two outputs in each IO in Table 3 will be emphasized as survey targets. First, some criteria are to be set for ensuring more objective judgment as to the degree of importance of survey targets and practicability of implementing surveys. In this study, two criteria were set for selecting outputs in light of the evaluation purpose: 1) an area of activities that directly contributes to the realization of the EO; and 2) an area in which donors have actively implemented or funded their projects. Also, another two criteria were set for selecting outputs in relation to the feasibility of surveys: 3) an area in which an external evaluator may easily conduct surveys; and 4) an area in which existing reports and data may be obtained. As an example, Table 5 shows the results of rating the outputs in IO1 applying these criteria. Finally, Outputs 1 and 2 in IO1, Output 4 in IO2, Output 1 in IO3 and Output 1 in IO4 were identified as core outputs. Next, core activities are to be identified. As shown in Table 3, activities c and h are included in a large number of projects in Outputs 1 and 2 in IO1, activity D in Output 4 in IO2, activity I in Output 1 in IO3, and activity O in Output 1 in IO4. These are core activities which will be emphasized in conducting analyses.

4.2.2 Considering Ways to Reduce the Number of Survey Targets and Identifying Data Collection Sites

The number of actual targets in field surveys is to be limited (Bamberger *et al.*2004). In this study, data collection sites in field surveys are considered after identifying core projects in each IO and their actual sites. Regarding IO1, for instance, projects no. 3, 5, 9 and 16 in Table 2 cover a wider variety of activities and thus, are identified as core projects. In project no. 9, activities were conducted mainly at four PPPGs, while in project no. 5, the main activity sites include 30 LPTKs (see notes in Table 3). In projects no. 3 and 16, activities were conducted mainly at three former IKIPs. These are potential data collection sites. Table 6 shows the core projects in each IO, and necessary data items for measuring the degree of realization of the EO and IO1-4 as well as the performance of core activities in each IO.

4.2.3 Considering Data Collection Methods and Minimizing Cost and Time

Considering data collection methods is important in order to minimize cost and time in field surveys (Bamberger *et al.*2004). Regarding the necessary data items shown in Table 6, any existing statistical data may be used for *A, D, E, O, P, U, V, W* and *X*, and thus, information sources may include Indonesia's Central Statistic Bureau (CSB), MoNE, and international institutions in education. It may be

Table 5 Results of Rating Outputs in IO1 Based on Four Criteria

	Criteria 1	Criteria 2	Criteria 3	Criteria 4	Total Points
Output 1	○	◎	◎	◎	11
Output 2	○	◎	◎	◎	11
Output 3	○	○	△	△	6

Source: The authors
Note: ◎ highly met (three points); ○ moderately met (two points); and △ minimally met (one point)

Table 6 Main Target Projects and Necessary Data Items

	Main Target Projects	Necessary Data Items
EO	–	*A, B, C, D*
IO1	3, 5, 9, 16	*E, F, G, H, I, J, O*
IO2	7, 9, 11	*P, Q, T*
IO3	12	*U, V*
IO4	13	*W, X*

Source: The authors
Note: See Table 2 for the numbers in target projects and Table 4 for the alphabets in data items.

Table 7 Data Collection Sites or Targets and Necessary Data Items

Data Collection Sites or Targets	Necessary Data Items
GDPSE of MoNE	*B, C, D, E, O, P, Q, U, W, X*
CSB, Examination Center of DGPSE	*A, E, U, V*
PPPG Bandung	*F, H*
Former IKIP Bandung (Indonesia Univ. of Education)	*I, J*
WB	*F, H, I, J, T*
ADB	*F, H, T*
JICA Experts (dispatched to MoNE)	*B, C*

Source: The authors
Note: BPGs and PKGs are not selected as data collection sites (data item *G*), considering that at those institutions effects from donors' projects are considered indirect.

necessary to conduct focus group discussions or questionnaires and interview surveys regarding *B, C, H, J* and *Q* which will mainly involve qualitative information. In this study, however, these methods are not adopted considering the above-mentioned constraints. Instead, it was decided to consider any existing surveys (or secondary data) regarding the target items. For *F, G, I* and *T*, existing output data may be obtained by visiting target projects sites and donors in charge. However, since it is difficult to visit all the sites due to the constraints, a single PPPG and a former IKIP in Bandung were selected from among the potential data collection sites. Table 7 shows final data collection sites, targets and necessary data items. However, this does not mean that the targets are limited only to these items.

5. Implementing Evaluation Surveys and Analyzing Results

5.1 Evaluation Surveys

5.1.1 Reconfirmation of Program Theory Model by Policy-makers

To have the assumed program theory model reconfirmed by policy-makers is important in field surveys. The relevance of causal relationships between the factors observed in the model forms the premise for value judgments in evaluation analysis. As mentioned earlier, any sort of arbitrariness in forming a model should be excluded to the extent possible. Thus, it will be meaningful later to make sure that there is no objection from those in charge of policy-making. Also, this reconfirmation process will be effective when discussing the validity of evaluation designs and results. In this study, the authors requested two officials at the Sub-directorate for Planning at the Directorate General of Primary and Secondary Education (DGPSE) of MoNE to reconfirm the

assumed model, and obtained comments and advice, which were reflected in the final version (Table 3).

5.1.2 Data Types and Collection Results

Types of existing data required for evaluation of donors' assistance at the sector level mainly include: 1) basic data –e.g. background information regarding the target sector and country; 2) statistical data at the outcome level; 3) secondary data regarding target survey items; and 4) output data from the projects. Regarding the third category, possible secondary data should be identified which potentially provides baseline data on control groups as survey targets. In the case that there is no secondary data, baseline data can be constructed by use of recall of the parties involved, or by using non-equivalent control groups (Bamberger *et al.*2004). This data is separated from the four categories and classified as: 5) alternative data. A considerably wide range of data is utilized for evaluation at the sector level and thus, effective use of secondary data is key to quality analysis.

Considering the evaluation purpose, an interrupted time series model in impact evaluation methodologies' was applied and these various types of data were collected. Especially, statistical data at the outcome level was emphasized in order to annually compare numerical values for each indicator, for the purpose of objectively grasping what changes were brought as to each target survey item during the period from 1990 until January 2001 when the decentralization policy became effective in the country. Table 8 shows the contents and types of data collected in field surveys regarding each necessary data item. It was decided to conduct analyses of each IO first, using these collected data, then to make value judgments regarding the realization of the EO from a holistic viewpoint. A summary of analysis and evaluation results is presented in the following sections.

5.2 Analyzing Survey Results

5.2.1 Financial State of Education Administration (IO2)

Data regarding the amount of capital input by

Table 8 Contents and Types of Data Collected in Field Surveys

Necessary Data Items	Contents of Collected Data (Data Types)
A	- Analysis results of changes in NEM at target schools in projects no. 2, 6 and 10 (3) - NEM at secondary education levels (2) - International comparative results of students' academic skills in science and maths education (2, 3)
D, W	- Numbers of students per teacher at each education level (1, 5)
E	- Percentages of in-service teachers meeting the set qualification at each level (1, 2)
F, I	- Numbers of in-service teachers who received re-training in target projects (4) - Numbers of in-service teachers whose academic degrees are upgraded in target projects (2)
O	- Numbers of teachers at each level (1)
P	- Budgets and expenditures of MoNE and DGPSE for each fiscal year (1) - International comparative results of governments' per capita education budgets (1)
T	- Quantities of educational materials distributed in target projects (4)
U	- Drop-out rates at each education level (1, 2) - Influence on school enrollment by the economic crisis occurring in 1997 (3)
V	- Gross enrollment rates at each level (1, 2) - Numbers of recipients of scholarships in target peojects (4) - Influence on school enrollment by the economic crisis occurring in 1997 (3)
X	- Numbers of schools at each level (1) - Numbers of school buildings/classrooms newly built or restored in target projects (4)
B, C, H, J, Q	(not collected)

Source: The authors
Note: Data types include: 1. basic data; 2. statistical data at the outcome level; 3. secondary data; 4. output data; and 5. alternative data.

the donor agencies in the target sub-sector will be presented in comparison with the education budgets and expenditures in Indonesia. The government's expenditures in PSE steadily increased from FY1990 to FY1996 (ADB1997, JBIC2002). DGPSE's development expenditure, besides the recurrent expenditure including salaries for the teachers, was calculated at about 200 million US dollars in FY1990 and increased to 790 million US dollars in FY1996, almost fourfold. The total amount including the recurrent expenditure in FY1996 was about 1,580 million US dollars. However, due to the economic crisis occurring in 1997, DGPSE's development budget decreased to about 100 million US dollars in FY1998, and recovered to about 500 millions US dollars by FY2000[8]. Meanwhile, the total amount of input in the projects implemented or funded by donors during the decade starting in 1990, as shown in Table 2, was calculated at about 1,650 million US dollars and thus, the average per year is about 165 million US dollars[9]. That is, the amount of input by donors accounts for quite a large portion of the total government budget in PSE[10].

5.2.2 Quality of In-service Teachers (IO1)

The development of educational personnel is an important factor in IO1, since they will directly or indirectly influence students, the final beneficiaries. In almost all the projects shown in Table 2, re-training or special courses were provided to in-service teachers, teacher trainers, laboratory technicians, librarians and management staff at the teacher education institutions in order to upgrade their academic degrees to the necessary levels. In-service teachers mainly participated in these re-training courses at PPPGs and former IKIPs, while the other personnel did so at various institutions, including some universities in developed countries.

By 1990, the lack of sufficient in-service teachers as well as their insufficient degree levels became an issue. Since 1998, Indonesia's national development plans, including REPELITA V, VI and PROPENAS, have set the academic levels of D2, D3 and S1 as necessary qualifications for teachers at the PR, JS and SS levels respectively[11]. In 1992, there were 1,154,000 primary, 375,000 junior secondary and 203,000 senior secondary teachers. 80 percent of all primary teachers, and more than 60 percent of junior and senior secondary teachers did not meet the qualifications (JICA1997, WB1996a and 1996b). By 2002, the numbers of teachers increased to 1,235,000, 467,000 and 378,000 respectively, and more than half of all primary teachers, and two-thirds of junior and senior secondary teachers had met the qualifications (MoNE2003). The net increase in the total number of teachers during the 10 years can be simply calculated at about 350,000, and the number of teachers meeting the qualifications as about 720,000. Meanwhile, in the target projects listed in activities c and h in Table 3, re-training courses were provided to approximately 150,000 in-service teachers at all target education levels. Among them, at least 70,000 upgraded their academic degrees to the necessary levels[12]. This number accounts for almost half of the total beneficiaries of the projects, and also about 10 percent of the net increase.

5.2.3 School Environment (IO4)

Regarding the school environment of students, the lack of the number of school facilities at all target education levels had become an issue by 1990, and MoNE made plans to increase their numbers. In 1992, there were 159,000 primary, 18,000 junior and 7,300 senior secondary schools, and by 2002 the numbers had increased to 171,000, 31,000 and 12,000 at each level (JICA1994b, MoNE2001). The net increase in the number of schools during the period can be simply calculated at about 30,000. Meanwhile, in the target projects listed in activity O in Table 3, approximately 1,450 school buildings and 3,500 classrooms were newly built and 2,650 school buildings and 5,400 classrooms were restored[13]. All the new school buildings were at the JS level. The number of newly built school facilities accounts for approximately 15 percent of the net increase.

5.2.4 Universal Educational Opportunity (IO3)

The gross enrollment rates[14] in FY1991 were 101.1, 52.7 and 36.7 percent at each education level, and those in FY2000 were 112.9, 73.0 and 45.0 respectively (JICA1994a, MoNE2001). While all the rates at three levels increased, there was an especially large increase in the rate at the JS level. The drop-out rates in FY1990 were 4.3, 6.3 and 3.8 percent at each level, and those in FY2000 were 2.6, 3.6 and 2.8 respectively (JICA1994a, MoNE2003). While the rates at all three levels decreased, similar to the enrollment rates, there was an especially large decrease in the rate at the JS level. Due to the economic crisis occurring in 1997, there were concerns at that time regarding negative influence on students' enrollment. Sakuma (2001) mentions that: 1) an increase in the drop-out rates at all three levels was clearly observed right after the crisis. However in the following fiscal year the rates decreased to the same level as before the crisis; 2) there was a relatively high decrease in the enrollment rate at the JS level in FY1997. However it was only less than one percent compared to FY1996 and then increased again at all three levels in FY1998. As a result, the influence on the entire target sub-sector was not as negative as expected, and rather transitory due to the government's intervention.

A total of more than 2 million students were provided with scholarships in the target projects listed in activity I in Table 3[15]. Considering the facts that all these projects were launched within two years before or after the crisis and that priority was given, in selecting candidates for scholarships, to those from poor families who were considered the most vulnerable to the crisis, the degree of contribution by the donors was fairly large in minimizing negative influence on students' enrollment.

5.2.5 Quality of Education (EO)

Considering the limited types of collected data, two different indicators are applied to observe the degree of realization of the EO. Firstly, in order to examine students' academic skills, NEM, the scores in former EBTANAS (or current UAN), which is a national standard examination held at the time of graduation at all junior and senior secondary schools, is applied as an indicator. According to one of the personnel at the Examination Center of DGPSE, it is possible to compare the average scores for each of the target subjects in different fiscal years[16]. Based on the data collected at the Center, there is no evident trend of increase or decrease observed in the average scores for any of the target subjects from FY1994 to FY2003.

Another set of analysis results regarding NEM data was conducted individually by the WB and ADB in projects no. 2, 6 and 10 in Table 2. Considering the fact that it is virtually impossible to access all schools that may be directly or indirectly affected by all the listed projects in Table 2, these results may be considered as showing representative data, since these projects implemented the same types of activities, such as c, q, z, D, G and N in Table 3, including some of the core activities, and thus may be considered typical in the target sub-sector. The final evaluation report of project no. 10 mentions that there was a slight increase in the overall average NEM at target schools before and after the project implementation. Regarding project no. 6, analysis results showed that there was a drastic increase in the overall average NEM in some provinces while there was no tendency of increase in other provinces. Regarding project no. 2, although the analysis results showed that there was an increase in the average NEM, the evaluator mentions that the results do not necessarily reflect the influence from the project. In other words, the actual situations seem to differ from project to project.

Secondly, in order to see the quality of education, the ratio of students per teacher is applied as an indicator. In FY1992 the ratios were 23.0, 14.5 and 11.7 at the PR, JS and SS levels respectively, and in FY2000 the ratios were 23, 16 and 14 (JICA1994b and 1997, MoNE2001). Although there was an increase at both the secondary levels, the overall ratios may be considered relatively low compared to other countries.

5.3 Summary of Evaluation Results

Conclusions derived from these analysis results are as follows: 1) in general, during the target period the Indonesian government largely depended on donors for funding, creating an environment in which donors' assistance projects would produce substantial influence in the target sub-sector; 2) there was an overall improvement in PSE regarding different aspects, including the quantity and quality of in-service teachers, the school environment of students, better opportunities for education, etc. The tendency of improvement was remarkable at the JS level, reflecting the goal that junior secondary education would be compulsory (see later), set by the government in REPELITA VI in 1994. An important part of the situational change in the entire sub-sector can be obviously attributed to donors' assistance projects, according to the data shown in previous sections; and 3) no clear indication, in the meantime, is observed as to the improvement at the EO level during the time period, using only the two applied indicators and the collected data. Therefore, the effects from donors' assistance projects appeared to be limited to only the IO level, based on the collected data. In other words, other types of indicators and data are needed in order to measure and examine the degree of realization of the EO in more detail, or that it may simply take more time for the impact to appear at the EO level.

5.4 Consideration of Evaluation Results

It is important to examine evaluation results from a holistic viewpoint considering the status of PSE at the end of the target period. This helps to observe the degree of donors' contribution to the development of the target sub-sector in a more relative manner. Since the enrollment rate at the primary level had increased almost 100 percent by 1993, the government set two important goals in its second 25-year development plans 1994-2018: 1) to realize a nine-year compulsory education system by increasing the enrollment rate at the JS level to 100 percent by 2008, and; 2) to improve the quality of in-service teachers by increasing, to more than 80 percent, the percentage of those whose academic degrees are upgraded to D2 and S1 at the primary and secondary levels. The data presented in the previous sections shows that the country was on its way to achieving these goals and that donors played an important role in contributing to the achievement during the target period.

However, according to international statistical data, Indonesia, at the end of the target period, remained one of the lowest countries in Asia in the quality of education systems, government's per capita education budgets, accessibility to education, and students' academic skills in science and mathematics at the PSE levels (JBIC2002, MoNE2003). After the decentralization policy became effective in January 2001, it is pointed out that the discrepancies in the quality of education are still large among regions (JICA2001). Although some donors have initiated their projects in a new decentralized approach, there are still areas in which donors can contribute.

6. Identifying Threats to the Validity and Adequacy of Evaluation, and Proposing Solutions

After considering evaluation results, threats to the validity and adequacy of the evaluation designs and results are to be identified (Bamberger *et al.*2004). In regard to the former, major threats are as follows.

6.1 Validity and Adequacy of Evaluation Designs

6.1.1 Program Theory Model

The issue of exclusion of arbitrariness was emphasized when defining the program theory model, however completely excluding arbitrariness is never an easy task. Also, only certain outputs and activities in each IO were considered in the analysis on the presumption that it is virtually impossible to conduct field surveys on each of the relevant activities due to their wide varieties in addition to the various constraints. As a result, there may be some limitation in the validity of the causal relationships on which the

analysis results are based. A solution is to further elaborate the means of setting criteria for selecting core outputs and activities in analysis, along with spending a considerable amount of time involving policy-makers when identifying a program theory model.

6.1.2 Interrupted Time Series Model

In analysis using this model, one has to observe the changes in a target group or society before and after an intervention, considering any possible influence by other exogenous factors (Ryu and Sasaki 2000). In this study, however, it is not easy to see "net impact" of donors' assistance projects on PSE for the following reasons: 1) donors' intervention was continuous; and 2) the government's intervention was also taking place simultaneously during the set time period. The method applied in this study, therefore, was to compare the situations between 1990 and 2000 (or equivalent years depending on whether necessary data is obtained) using numerical values and to make judgments on the attribution to donors' assistance during the period. A solution is to devote much time and energy to obtaining necessary data within a realistic range in order to improve the accuracy of analysis.

6.1.3 Setting Evaluation Targets and Distinguishing Beneficiaries

In the evaluation purpose, clear distinctions among beneficiary groups were not made between students: 1) at the PR, JS and SS levels; 2) at public and private schools; 3) at schools under MoNE and schools under MoRA; 4) in different areas of subjects or courses; and 5) in different regions. The reason for this is that donors' assistance was provided in more than one classification at the same time in many target projects. It was also because of the authors' judgment on the applicability of the program theory model based on a common basic structure of the causal relationships, regardless of these classifications. As a result, analysis results were presented at a rather macro level, focusing on the EO and IOs. It is, however, possible to make those distinctions through analysis using different models under new evaluation purposes, especially emphasizing different factors among the classifications, so that data may be interpreted from different viewpoints and at a more micro level.

6.2 Validity and Adequacy of Evaluation Results

This issue needs to be discussed in relation to the observations through data collection.

6.2.1 Data Types and Characteristics

Impact evaluation at the sector level requires the use of different types of data, including secondary data and statistical data at the outcome level. Utilizing more relevant existing data, on the presumption that data is trustworthy, is key to improving the depth of analysis. However, in most cases, there is no way to examine the accuracy or credibility of data, regardless of whichever institution compiled and publicized it. This may be the case, especially when statistical data at the sector level is drawn from the recipient country where data collection systems or infrastructure for such may not be well equipped, and when only survey results are presented in secondary data. A solution is to request those institutions in charge to provide relevant information and clarification for a better determination of data credibility.

6.2.2 Data Availability

Beyond the fact that collecting expected data was sometimes difficult due to various constraints, some types of data seemed not to exist, although they were necessary for grasping the status of donors' assistance or some important elements in the sub-sector. In this study, not all necessary data was collected, including data regarding items involving qualitative information, as shown in Table 8. This may likely hinder the degree of depth in analysis, especially regarding the realization of the EO and its attribution to donors' assistance. A measure taken in this study was to apply some equivalent data for uncollected data - e.g. the enrollment rates in FY1991 instead of those

in FY1990, although this may affect the consistency and accuracy in the analysis results. The limitation in the availability of data is partly due to donors' roles and responsibilities (see later).

6.2.3 Data Interpretation

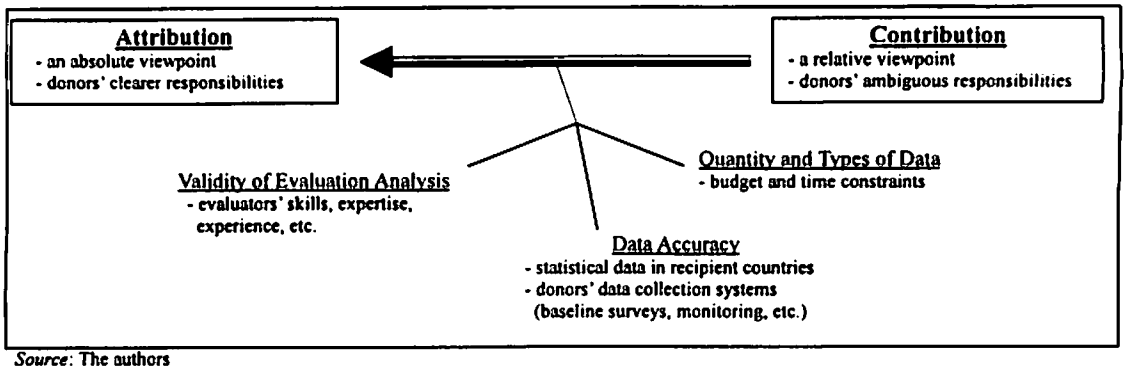
Deciding what types of indicators are to be applied and how to interpret collected data is always a difficult issue, especially when it comes to measuring education quality. In this study, many indicators were drawn from MoNE's documents, and some of the goals raised in their education policies were applied as standard criteria for judgments, from a relative viewpoint, on the degree of contribution of donors' assistance in PSE. However, measuring the impact on the realization of the EO and IOs as well as their attribution to donors' assistance from an absolute viewpoint is extremely difficult, especially in the situation where not enough data is collected for a particular indicator. Another issue to be considered is that impact on education quality may not be grasped correctly if the development of students is also not observed and analyzed at a micro level.

7. Summary and Prospects

This paper presented, as a case study, a framework for impact evaluation of donors' assistance projects at the sector level. The main differences between targeting only a single project and multiple projects in impact evaluation include: 1) the volume of information to be dealt with. Sorting out and ordering the information is key to using it in a more efficient manner; 2) a wide coverage of areas, sometimes crosslink with different sectors; and 3) evaluation viewpoints emphasizing a realization of outcomes at higher levels. In regard to the issue of the limitation in the availability of data, the authors were able to find only a few project documents in which donors presented some original outcome data mainly collected at their project sites. In those cases, however, they did not necessarily present in detail any survey method applied or grounds on which the conclusion or results were based. Moreover, in most cases, the documents presented no baseline data, simply using relevant (or sometimes indirectly relevant) statistical data published by the recipient country's government. Considering the situation, this may be an area of improvement which donors should consider.

Regarding the issue of "attribution" and "contribution" in evaluation, Figure 1 shows a conceptual model of the differences between the two. It is easier to observe donors' contribution to the development of a target sector than to observe its attribution to their projects, because: 1) contribution can be observed from a relative viewpoint (e.g. comparison among donors), while attribution from an

Figure 1 Conceptual Differences between "Contribution" and "Attribution" in Evaluation



Source: The authors

absolute viewpoint; 2) a larger volume and variation of data regarding relevant research items are required to observe its attribution; and 3) more accuracy in collected data and 4) validity of evaluation analysis are key to ascertaining the degree of attribution.

This framework can be applied to other sectors in any country where donors implement assistance projects at a certain level of intensity. Also, since results derived from evaluation using this framework may cover a wide range of issues of the target sector over a long term, it will be useful for making recommendations in the formulation of the government's overall policy in the sector, as well as donors' aid policy.

Acknowledgement

My deepest gratitude goes to Prof. Koichi Miyoshi at Ritsumeikan Asia Pacific University who provided useful comments in the preparation of this study, as well as to the reviewers and everyone who cooperated in this study.

Notes

1 The reasons for setting the target and purpose are as follows: 1) after the World Conference on Education for All in 1990, an international focus on the importance of assistance in basic education was established; and 2) several donors have implemented or funded numerous projects in PSE in Indonesia since then.

2 Following the evaluation purpose, this study does not deal with other projects launched under new decentralized administrative systems after January 2001.

3 Refer to Miyoshi et al. (2003) for program theory and the PTM.

4 In REPELITA V (1989-93), VI (1994-98) and PROPENAS (2000-04), the same major issues were raised regarding PSE, such as the necessity of improvement in the enrollment rates, the quality of in-service teachers, school environment, textbooks and curriculums, educational administration, etc.

5 In Indonesia, primary education refers to six-year elementary school, and secondary education refers to three-year junior high school as well as three-year senior high school.

6 Table 2 includes projects in which the Ministry of Religious Affairs (MoRA) is involved (see later).

7 Refer to Ryu and Sasaki (2000) for the details, for instance.

8 MoRA's education budgets are not incorporated because, based on the authors' calculation, its development expenditure was less than one-tenth of that of MoNE and thus, considered relatively small enough to ignore.

9 These numbers are calculated in an approximate manner.

10 According to the authors, donors' total input in PSE from FY1990 to 2000 accounted for the average of about 20 percent of DGPSE's total annual budgets including the salaries for teachers every year.

11 D2, D3 and S1 refer to the levels of graduates from a two-year and a three-year college, and a university respectively (JICA2001).

12, 13, 15 These numbers are calculated based on data in relevant documents to the target projects.

14 The gross enrollment rate refers to the percentage of the total number of students enrolled at a certain level of education compared to a certain school age-group population. School age-group populations at the primary, junior and senior secondary levels include those at 7-12, 13-15 and 16-18 years of age respectively. At the primary level, the percentage is above 100 percent because: 1) some are already enrolled before the age of seven years; and 2) others are still enrolled after the age of 12 (JICA1994a).

16 Through the application of the item response theory, the Center implements a trial test by sampling a certain number of students at each education level, calculates the difficulty for the correct answer for each question item, and selects, for the nation-wide examination, only questions whose difficulties are close to the average value. Appropriate sampling is the premise for the validity of this method.

References

Asian Development Bank (1997). *Report and*

Recommendation of the President to the Board of Directors on Proposed Loans to the Republic of Indonesia for the Second Junior Secondary Education Project.

Bamberger, Michael, Jim Rugh, Mary Church, and Lucia Fort (2004). *Shoestring Evaluation: Designing Impact Evaluations under Budget, Time and Data Constraints.* The American Journal of Evaluation, 25(1), 5-37.

International development Journal (2004). *Lexicon of International Cooperation.* 3rd Edition.

Japan Bank for International Cooperation (2002). *The Current Status and Issues in the Education Sector.*

Japan International Cooperation Agency (1994a). *Report by the Committee for Country-based Assistance for Indonesia.* 2nd Edition.

---- (1997). *Ex-ante Evaluation Report of Project for Development of Science and Mathematics Teaching in Primary and Secondary Education.*

---- (1994b). *Sector-based Basic Information of Indonesia,* Vol. 4.

---- (2001). *Sector Issues-based Basic Information of Indonesia,* Vol. 1.

Ministry of Foreign Affairs of Japan (2003). *Evaluation Report of Economic Cooperation 2003.*

Ministry of National Education of Indonesia (2003). *Educational Indicators in Indonesia 2002/2003.*

---- (2001). *Indonesia Education Statistics in Brief 2000/2001.*

Miyoshi, Koichi, Satoshi Morita, and Yoshio Aizawa (2003). *Toward Constructing More Suitable Program Theory for Japan's Evaluation.* The Japanese Journal of Evaluation Studies, 3(2), 40-56.

Ryu, Yoshiaki, and Ryo Sasaki (2000). *Evaluation Theories and Techniques.* Tagashuppan.

Sakuma, Jun (2001). *Study on the Impact of Economic Crisis in Indonesia on Primary and Secondary Education and the Effect of Social Safety Net programs by the Government of Indonesia.* Studies for International Cooperation, 17 (1), 47-55.

The World Bank (1996a). *Staff Appraisal Report for Junior Secondary Education Project in Central Indonesia.*

---- (1996b). *Staff Appraisal Report for Secondary School Teacher Development Project in Indonesia.*

(2006.7.31受理)