Political Implications from Empirical Analysis of the Performance-Based Evaluation System in National R&D Programs

Seung-Tai Kim[1], Seong-Jin Kim[1], Ki-Jong Lee[1*]

Abstract

This study was conducted to draw political implications from an empirical analysis of performance-based evaluation systems of current National R&D Programs. For this, major issues on existing evaluation systems were identified through tracing and analyzing transition processes, investigating and analyzing related literature, and knowledge learned through experiences. Questionnaire research on how these issues are recognized by personnel concerned with R&D evaluation was conducted through empirical analysis. The result showed that the triennial evaluation system conducted from 2009 was generally recognized as helpful in alleviating evaluation pressures; however, the evaluated subject was still perceived poorly in terms of importance of R&D program evaluation and experience pressures in making evaluation data. Furthermore, there is plenty of room for improvement since evaluation does not reflect the characteristics of R&D activities and the result is not fully utilized for internal purposes.

1. Introduction

In many advanced countries, a large portion of investment from the government investment budget is focused on supporting the development of science and technology research capable of producing a significant economic ripple effect as a basis for creating knowledge and solving social problems. In Korea, there has been a constant effort since the 90s to expand the size of investment by nearly 10% to raise the R&D investment level to that of of advance countries. In 2009, it allotted a substantial budget of KRW12.3437 trillion – an 11.4% increase from the previous year (Comprehensive Guideline to National R&D Program, 2009). The size of government investment in R&D is expected to continue to rise in the future under the policies of the current government including the recommendations of the '577 Initiative' – a plan to expand investment by 5% of GDP with the goal to rank among the top seven science powers by 2012 – as part of the Science & Technology Basic Plan prepared by Myung-Bak Lee government.

With this large expansion in the R&D budget, the focus has shifted from the management of input and output to an emphasis on expanding performance-based investment efficiency. According to the Lee administration's Science & Technology Basic Plan, discussion has begun on system improvement plans to alleviate evaluation pressures of researchers by suggesting the "Establishment of Researcher Friendly R&D Management and Evaluation System" as a key task.

However, studies on the main factors that pressure researchers are still insufficient as we begin to go

[1]Office of S&T Policy & Planning, Korea Institute of Science & Technology Evaluation and Planning(KISTEP), Seoul, 137-130, Korea
[*]Corresponding author. E-mail: 27ljong@kistep.re.kr

forward with these efforts to improve the system. In relations to institutional evaluation programs, Chan-Gu, Lee (2009) only presented the problems of current systems and system improvement plans in his in-depth interview. Existing studies failed to provide empirical verification through quantitative methods on the impact of government evaluation systems to actual researchers after legislating "Law on Performance Evaluation and Management of National Research and Development Projects". Therefore, it is necessary to understand how the evaluation system is perceived by relevant personnel and identify future improvement directions for gradual improvement of the evaluation system. The objective of this study is to examine how improvement of the evaluation system on national R&D activities is embraced in the field and in which direction improvement is required in the future through conducting and analyzing questionnaire research to evaluation personnel to present the problems of the new evaluation systems and suggest improvement plans.

This study analyzed the evaluation system to develop methodologies for measuring the effectiveness of the improved evaluation systems and, simultaneously, investigated and analyzed various evaluation related releases and dissertations on evaluation theories to highlight major issues relevant to existing R&D evaluation systems.

To investigate recognition by relevant personnel of major issues, a questionnaire was designed and its results analyzed. The normalization process was pursued through a discussion with experts regarding the major issues for detailed questionnaire questions and incorporating them into the general evaluation systems framework. Meanwhile, the questionnaire was conducted by categorizing evaluation personnel into evaluator and evaluation subjects to study current status of recognition as well as the difference of recognition between the parties. Finally, the results analyzed to ultimately draw political implications.

2. National R&D Program Evaluation System

2.1 National R&D Program Evaluation System

2.1.1 Transition Process of National R&D Program Evaluation Systems (Table 1)

Before legislating the "Special Act on Innovation in Science and Technology" (Apr. 10, 1997, Law No. 5340), the R&D projects were evaluated with other programs. With the legislation of the Special Act in 1997, the importance of R&D management was stressed and specialized evaluations on R&D programs

Table 1 Transition Process of National R&D Program Evaluation System

Type	1998~2000	2001~2004	2005~2007	2008~2009
Improvement Details	• Promotion in status from bureau to Department of Science and Technology (Feb. 98) • Installation of National Science and Technology Council (Mar. 99) • Establishment of '5-Year Science and Technology Innovation Plan' (Dec. 97) • Enactment of 'Special Act on Innovation in Science and Technology' (Apr. 98)	• Enactment of 'Basic Law on Science and Technology (Jul. 01) • Establishment of the 1st Science and Technology Basic Plan (May. 03)	• Implementation of system of Deputy Prime Minister of former Department of Science and Technology (Oct. 04) • Installation of Science Technology Innovation Division in former NSTC (Oct. 04) • Enactment of 'Law on Performance Evaluation in R&D Programs' (Dec. 05)	• Transfer of colligated authorities of evaluation planning from Ministry of Education to Ministry of Strategy and Finance through government restructuring (Mar. 08) • Establishment of the 2nd Science and Technology Basic Plan (Aug. 08)
Significance	Implementation of categorized evaluation on national research development program	Preparation of comprehensive coordination basis for R&D activities	Reinforcement of comprehensive coordination function and implementation of performance-based evaluation	Alleviation of evaluation pressures and reinforcement of evaluation efficacy
Emphasis	Prevention of duplicative investment in R&D activities	Validity evaluation of financial input and execution	Establishment of performance-based management systems	Implementation of triennial evaluation and in-depth evaluation

undertaken for the first time. The main focus of the R&D program evaluation at that time was on finding and arbitrating repeated programs or those that required relation among the programs. Furthermore, the basic framework of science and technology policy was constituted through establishing a '5-Year Science Technology Innovation Plan (1998-2002)' for the first time that incorporated mid to long term planning for science and technology and inaugurated the National Science & Technology Council (NSTC), the highest decision-making organization regarding national science and technology policies, in January 1999. At the same time, the former Science and Technology Bureau was elevated to the status of Department of Science and Technology in February 1998, attesting to the rise in importance of science and technology.

The most important changes in the current execution and management system of science and technology were presented through enactment of the "Basic Law on Science and Technology" (Jan. 16, 2001, Law No. 9089) (the 'S&T Basic Law'). The system of the 'Science & Technology Basic Plan' (1st: 2003-2007; 2nd: 2008-2012) was prepared through the legislation of subordinate laws to the Special Act, which also prepared the opportunity to establish and execute science and technology policies by identifying regulations for the 'Comprehensive Promotion Plan for Regional Science and Technology' and items regarding NSTC. Moreover, it established the basis for a national science and technology innovation system, including research development, and investment and human resources to ensure comprehensive and long-term development of science and technology. From this period, the evaluation system went further to consider the feasibility of budget input and execution process beyond the level of simple repeated investment arbitration along the expanded investment sizes. However, evaluation did not pay much attention to R&D performance until the legislation of the "Law on Performance Evaluation and Management of National Research and Development Programs (Dec. 30, 2005, Law No. 8852)" (the 'Performance Evaluation Law'). This law identified efficient management systems for evaluation and performance of R&D activities and divided the existing two categories of contracting institution evaluation and R&D business evaluation into self and meta evaluation and performance plan management and follow-up and specific evaluations and stipulated the basic principles of performance management for systematic management of the results. In addition, it identified the role of evaluation result for policy establishment, program execution and budget coordination and prepared the current performance-based evaluation system just as the most important event in the history of science and technology management system was the enactment of the 'S&T Basic Law', so too was the enactment of the 'Performance Evaluation Law' in the history of R&D activity evaluation. After this, many supplementing plans began to be proposed for the effective establishment of a performance-based evaluation system.

2.2 Current Status of 2009 National R&D Program Evaluation and Characteristics of System Improvement

2.2.1 2009 National R&D Program Evaluation System

Under the current evaluation system for the national R&D program, self-evaluation and performance management are conducted by each department and meta evaluation, follow-up evaluation and specific evaluation(in-depth approach) are conducted by the Ministry of Strategy and Finance. The characteristics of each evaluation program are as follows(shown in Figure 1).

First, for the self-evaluation system, the objective and details, execution systems, and performance evaluation systems are investigated considering whether the program created proper results according to the performance plan established by each department concerned. After the self-evaluation, meta evaluation is conducted on the result of the self-evaluation and evaluation system to judge whether the self-evaluation was conducted appropriately by the Ministry of Strategy and Finance. According to the result of meta evaluation, the evaluation result is confirmed for the final program and this confirmation is then used for establishing and revising budgetary and execution plans. At this point, the Ministry of Strategy and Finance, which plays comprehensive role, develops

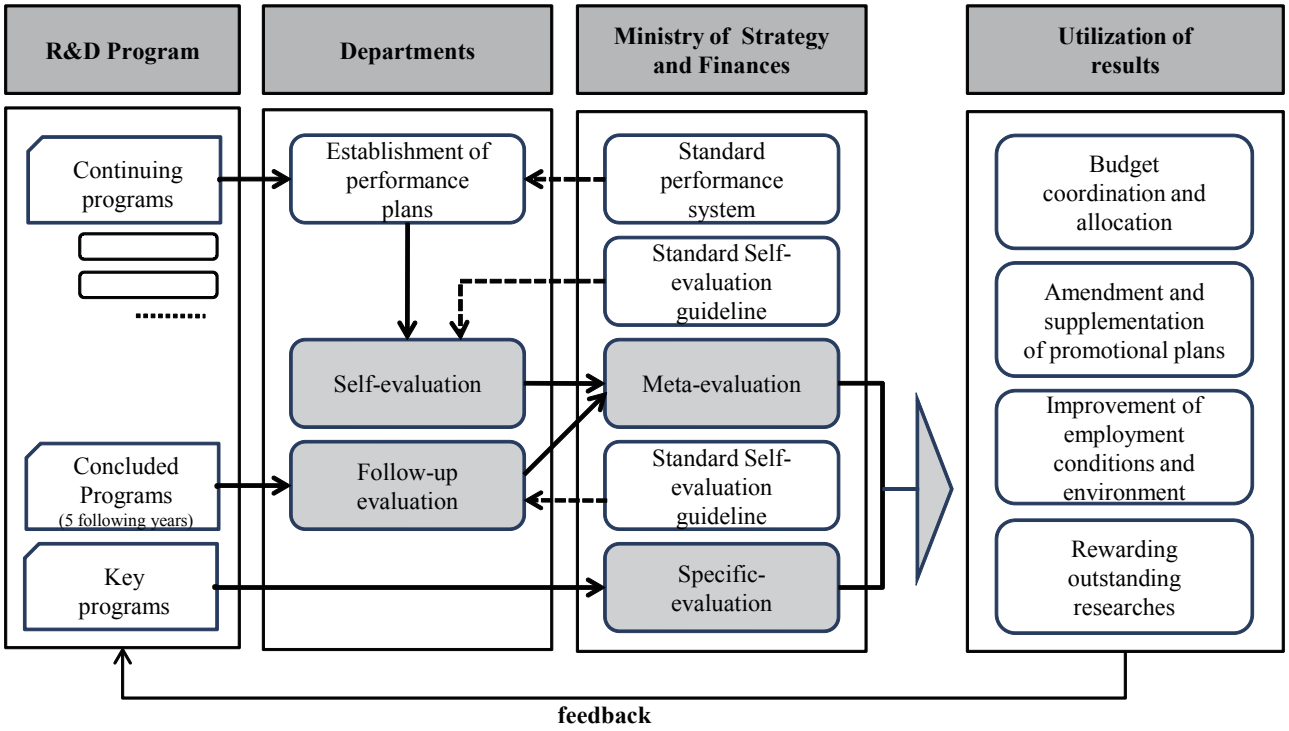

Figure 1 Transition Process of National R&D Program Evaluation System

Standard Performance Index Systems and Standard Evaluation Guidelines in support of the establishment of performance plans and providing unified directions in planning. In 2009, it conducted self and high ranking evaluation on 73 national R&D programs in 16 government ministries, including the Ministry of Education and Science, and the Ministry of Land, Transport and Maritime Affairs. This corresponds to about 1/3 of 207 national R&D programs in total (as of Jan. 2009).

The specific evaluation is hosted by the Ministry of Strategy and Finance for programs, including long-term and large sized programs, repetition arbitration and connection programs, multilateral departments' collaborative programs, and nationally and socially pending issued programs. In 2009, 10 programs and groups (8 ministries with programs totalling KRW820.4 billion) including a program to support an industry, university and research collaboration system, a program to enhance the material and components industry's competitiveness, and promote technology transfer commercialization groups, were selected and implemented.

Although follow-up evaluation is statutorily required, it is based on the evaluations of concluded programs. For this reason, it has never been executed due to issues of evaluation effectiveness. The Law stipulates a similar process for the self evaluation that the Ministry of Strategy and Finance develops and provides standard guidelines and each department conducts evaluation on this basis. But it evaluates only the programs within 5 years of conclusion and its focuses are on the perspective of management and utilization of the program's result.

2.2.2 Characteristics of 2009 National R&D Program Evaluation System Improvement Directions

The characteristics of the 2009 National R&D Program evaluation system improvement can be categorized into three: alleviation evaluation pressure on researchers, reinforcement of customized and consulting evaluation system through in-depth evaluation, and promotion of utilizing evaluation

results. Each category is specified below.

First, as part of effective reinforcement plans through creation of results, a measure to alleviate researcher's evaluation pressures was undertaken. Formerly, the researchers had to go through selection evaluation, annual evaluation, stage evaluation, and follow-up evaluation and only then did the programs go through self and meta evaluation, specific evaluation, performance plan confirmation and inspection. On top of this, self and meta evaluation of the contracting institutions was conducted in which the individual researcher annually was pressured with annual evaluation, self and meta evaluation of the program, and self-evaluation for the affiliating institution at a minimum.

Therefore, the evaluation cycle of program self and meta evaluation was extended from 1 year to 3 years and the yearly evaluation subjects were reduced by 1/3 in an effort to alleviate evaluation pressures on the researcher. Under the former system, a total of 207 programs from 17 ministries (KRW8.1489 trillion) were subject to evaluation (as of Jan, 09); whereas, with the extension of its cycle to 3 years, the subject was reduced to a total of 73 programs for 16 ministries (KRW2.2109 trillion). The re-evaluation system, which repeated the self-evaluation when the result of meta evaluation was proved invalid, was also abolished and revised as a form to draw final evaluation results of the program at the meta evaluation level. The program evaluation, which often occurred three times a year in the worst case, was improved in that evaluation occured once in three years. This improvement is expected to alleviate researcher's evaluation pressures by mandating research and program performance evaluation for contracting institutions once in three years.

Measures to reinforce evaluation effectiveness accompanied those alleviating researcher pressure. The most important improvement was to bring in-depth analysis to in-depth evaluation. In contrast to the existing specific evaluation method, which conducted the inspection through a commonly applied checklists, the specific evaluation through in-depth analysis differed in that it establishes separate evaluation strategies, conducts performance analysis, and produces evaluation results accordingly. That is, the in-depth approach evaluation is a method that reviews whether the establishment of evaluation strategies and creation of mid and long term program performance were produced appropriately. In brief, it is an in-depth evaluation.

There have been changes to reinforce the effectiveness of not only for the in-depth evaluation but also for self and meta evaluation. The past practice of constituting a technology specialist oriented evaluation council was changed to include at least one specialist from the areas of economy, humanity and society in order to have an overall perspective on program execution systems as well as program performance.

Lastly, to maximize the utilization of evaluation results, measures were taken to reinforce connectivity to the budget. The former method of utilizing budget allocation as a point of reference was transformed to preparing a concrete connection standard to enable a direct and realistic budget connection according to the evaluation result. In other words, it stipulates the principle of downsizing the budget by 10% for insufficient programs, freezing the budget for average programs, and expanding the budget for superior programs.

3. Empirical Analysis on the Recognition of Evaluators and Evaluation Subjects

The performance-based evaluation system for National R&D programs is introduced to establish a performance management and evaluation system that is distinct from simple results-oriented measurement in order to expand public sector investment efficacy within the context of the global commencement of a knowledge-based economic system in the 21st and the expansion of R&D investment (Sang-Yup Lee, 2007). However, there has been no substantial intermediary evaluation on the performance evaluation system itself thus far.

This study plans to identify future improvement factors through a quantitative analysis of how broad evaluation results affect research management centered on National R&D Program evaluation. In addition,

it will address the question whether the issue of expertise of evaluation counselors is an important topic through quantitative studies related to the expertise of evaluation counselors. It will also review issues related to evaluation systems such as the recognition of evaluators and evaluation subjects of current triennial evaluation systems through the questionnaire.

3.1 Investigation and Design of the Empirical Analysis Framework

There have been many preliminary studies on comprehensive evaluation systems and each individual element in evaluation system from academia. There also have been many formal or informal discussions on TF activities or evaluation sites although they were not publicized through provisional improvements or preceding studies. Assessing these two elements scientifically is essential to ensure the effectiveness of the performance-based evaluation system (Chan-Gu Lee, 2009).

As we have seen earlier, the government has continued its effort to improve the National R&D Programs evaluation system. The Related issues are presented and regulated through discussions with specialists to draft detailed questionnaire items.

Dong-Hoon Oh (2006) suggested a method of categorizing items relevant to the evaluation systems –evaluation philosophy, evaluator, evaluation organization, evaluation cost, evaluation process, evaluation system, evaluation management, result reflux plan, monitoring, related laws, and evaluation culture – into 5 categories: the evaluation paradigm, evaluation resources, evaluation activities, reflux systems, and evaluation environment. In this study, issues investigated using this framework were categorized and normalized(shown in Table 2).

Questionnaire items, for the questionnaire research were divided into evaluators and evaluation subject in recognition of supposed differences between the two groups. Table 3 shows derived questionnaire items.

The 5-Point Likert Scale was used to measure the questionnaire result and each variable was assessed through 4-6 multiple choice items (some short answer questions). From the Cronbach's Alpha Test, which measures reliability and validity of questionnaire questions for each concept, all of the above variables were above 0.7 showing a high level of reliability and validity of the questionnaire items created to measure each variable.

Table 2 Constituent of the Evaluation System

Constituent	Content
Evaluation Paradigm	Evaluation philosophy, objective, principle, subject, scope, etc
Evaluation Resources	Evaluator, evaluation organization, evaluation cost, etc
Evaluation Activity	Evaluation process, evaluation system, performance management, etc.
Reflux System	Reflux method, reflux subject, evaluation result monitoring
Evaluation Environment	Related Law, information system, evaluation culture, evaluator training, etc

※Dong-Hoon Oh (2006), 「Study on National R&D Evaluation System Establishment」

3.2 Composition of Questionnaire and Respondent Characteristics

The questionnaire was developed into a webpage by dividing the user into either evaluator or evaluation subject who were then sent emails with the URL through which the subject participates in the questionnaire. The term of research was 10 days from December 18 to 26, 2008 for the evaluation and 7 days from December 18 to 26, 2008 for the evaluation subject(Table 4). To raise the reliability of the questionnaire result, the questionnaire subject utilized the evaluation participation human resources information registered with NTIS – a government's National R&D Program management service – and selected from among the personnel that participated in the National R&D Program in 2008.

The evaluators were selected from personnel from each department and private evaluation counselors who participated in evaluation and management of R&D activities while the evaluation subjects were selected from personnel in national and public research centers, government subsidized research institutes, and university and private corporation institutions who were the subjects of program evaluations and task evaluations. A total of 680 personnel, including

Table 3 Derivation of Questionnaire Items for Recognition on Evaluation System

Constituent	Major Issues	Questionnaire Items	Subject	
			Evaluator	Subject
Paradigm	Disputes over recognition of level of importance on the evaluation system	Most important system among evaluation systems for R&D activities	O	O
		Evaluation system that imparts most pressures among evaluation systems	O	O
	Disputes over level of satisfaction on the evaluation system	Level of satisfaction of evaluation results	O	O
	Disputes over replicated evaluation systems	Experience of being subjected to multiple evaluations at the similar period	O	O
		Period of repeated evaluations	O	O
		Appropriateness of current evaluation periods and the time	O	O
	Disputes over replicated evaluation contents	Evaluation system which the evaluation content is thought to be repeated	O	O
		Necessity and method of data sharing between evaluation programs	O	-
		Degree of differences in contents and formats of evaluation data among evaluation systems	-	O
	Disputes over evaluation pressures	Degree of pressure of evaluation tasks on research tasks	O	O
		Most oppressive factors in conducting evaluations	O	O
		Time spent writing evaluation and data	O	O
Evaluation Resources	Disputes over expertise of evaluation counselors and fairness	Precedence between expertise and fairness of the evaluation counselor	O	O
		Level of expertise of the evaluation counselor	O	O
		Cause behind low profile expertise of evaluation counselors	O	O
		Level of fairness of the evaluation counselors	O	O
		Cause of unfairness of evaluation counselors	O	O
Evaluation Activities	Disputes on reflecting research characteristics	Level of evaluation methods in reflecting characteristics of R&D activities	O	O
		Cause of failure of evaluation methods in reflecting characteristics of R&D activities	O	O
		Level of desirableness of the verification management system after voluntary presentation of the performance plans	O	O
		Level of reflecting R&D activity characteristics in the performance plans	O	O
		Cause of failure of performance plans in reflecting R&D activity characteristics	O	O
Reflux System	Disputes over internal use of evaluation results	Level of researcher utilization of evaluation results	O	O
		Reliability and fairness of evaluation results	O	-
		Internal utilization contents of evaluation results	-	O
	Disputes over utilization of evaluation results	Opinions on integrating evaluation results in the budget for the following year	O	O
	Disputes over reliability of evaluation results	Experiences and causes of different evaluations results from different evaluation authorities	O	O
		Experiences and causes on different results on the data similar to the previous year	-	O
Environment	System improvement demands by concerned personnel	System improvement requirements (short answer)	O	O

Table 4 Ratio of Questionnaire Result Responses

Type		Evaluator	Evaluation Subject	Total
Respondent	N	110	570	680
	%	16	84	100

110 evaluators and 570 evaluation subjects responded during the research period. In the case of the evaluators, the questionnaire was requested by phone for a higher rate of response; whereas for evaluation subjects, it only analyzed the result of the collected questionnaires due to the large number of subjects.

Looking at the characteristics of the respondents, 47.2% were from metropolitan areas, 15.4% in Daejeon, and 37.3% in nonmetropolitan area. Age-wise, 18.4% were under 40 years old, 53.2% were between 40-50, 26.9% were between 50-60, and 1.5% were over 60. In regards to the affiliated institutions, 43.7% were from universities, 21.9% from corporations, 16.8% from government-subsidized research institutions, 6.0% from national and public research institutes, and 11.6% were from elsewhere. For duration of employment, 16.3% worked than 5 years, 20.6% worked between 5-10 years, 37.6% worked 10-20 years, and 25.4% worked more than 20 years. For the experienced evaluators, 55.0% participated in institutional evaluation, 32.4% participated in task evaluation and 12.7% participated in program evaluation (Table 5).

3.3 Questionnaire Result

3.3.1 Recognition on R&D Activity Evaluation Status (Table 6)

To the question on the importance of evaluation systems regarding R&D activities, the evaluators responded in the order of 'Task Evaluation (47.3%) and National R&D Program Self Evaluation (44.5%). Among the evaluation subjects (the 'subject'), the response, "task evaluation is most important" recorded the highest at 60.5% which was. This corresponds to the general recognition that evaluator involved in overall inspection of the National R&D Program would perceive the importance of self-evaluation higher. On the other hand, both evaluators and subject chose 'self-evaluation' as the most pressured evaluation (60.9% of the evaluator, 46.5% of the subject; same

Table 5 Respondent Characteristics

Type		Evaluator		Evaluation Subject		Total	
		N	%	N	%	N	%
Region	Metropolitan Area	80	72.7	241	42.3	321	47.2
	Daejeon	17	15.5	88	15.4	105	15.4
	Non-Metropolitan Area	13	11.8	241	42.3	254	37.3
Age	Under 40	59	53.6	66	11.6	125	18.4
	Between 40-50	41	37.3	321	56.3	362	53.2
	Between 50-60	9	8.2	174	30.5	183	26.9
	Over 60	1	0.9	9	1.6	10	1.5
Affiliated Institution	University	10	9.1	287	50.4	297	43.7
	Corporations (Research Institutes)	4	3.6	145	25.4	149	21.9
	Government Subsidized (RI)	23	20.9	91	16	114	16.8
	National and Public (RI)	11	10	30	5.3	41	6.0
	Others	62	56.4	17	3	79	11.6
Years of Employment	Less than 5 years	41	37.3	70	12.3	111	16.3
	Between 5-10 years	24	21.8	116	20.4	140	20.6
	Between 10-20 years	30	27.3	226	39.6	256	37.6
	More than 20 years	15	13.6	158	27.7	173	25.4
Evaluation Experience	Institution Evaluation	87	79.1	511	89.6	598	55.0
	Task Evaluation	46	41.8	306	53.7	352	32.4
	R&D Program Evaluation	34	30.9	104	18.2	138	12.7

Table 6 Recognition on R&D Activities and Government Subsidized Research Institutions Evaluation Status

Questions	Subject	Responses									
		Item1		Item 2		Item 3		Item 4		Item 5	
Most important evaluation	Evaluator	Task Evaluation	47.3	Program Evaluation	44.5	Institution Evaluation	8.2	-	-	-	-
	Subject	Task Evaluation	60.5	Program Evaluation	31.4	Institution Evaluation	8.1	-	-	-	-
Most burdensome evaluation	Evaluator	Task Evaluation	13.6	Program Evaluation	60.9	Institution Evaluation	25.5	-	-	-	-
	Subject	Task Evaluation	21.2	Program Evaluation	46.5	Institution Evaluation	32.3	-	-	-	-
Level of satisfaction over evaluation results	Evaluator	Very high	1.8	High	30.9	Average	42.7	Low	22.7	Very low	1.8
	Subject	Very high	3.0	High	39.6	Average	42.3	Low	13.2	Very low	1.9
Experience and time of being subjected to multiple evaluations during similar period	Subject (N=182)	Yes	31.9	None	68.1	-	-	-	-	-	-
		1st Quarter	21.4	2nd Quarter	20.3	3rd Quarter	12.1	4th Quarter	46.2	-	-
Appropriateness of current evaluation periods and time	Evaluator (N=24)	Highly appropriate	0.9	Appropriate	27.3	Average	50.0	Inappropriate	20.0	Highly inappropriate	1.8
		1Quarter	29.2	2Quarter	45.8	3Quarter	8.3	4Quarter	16.7	-	-
	Subject (N=65)	Highly appropriate	1.2	Appropriate	29.6	Average	57.7	Inappropriate	9.8	Highly inappropriate	1.6
		1Quarter	33.8	2Quarter	20.0	3Quarter	33.8	4Quarter	12.3	-	-
Evaluation system which the evaluation content is thought to be repeated	Evaluator	Task/ Program	44.5	Task/ Institution	5.5	Task/Task	5.5	Program/ Institution	22.7	Not repeated	21.8
	Subject	Task/ Program	43.3	Task/ Institution	10.4	Task/Task	10.0	Program/ Institution	8.4	Not repeated	27.9
Necessity of data sharing between evaluation programs	Evaluator	Highly necessary	12.7	Necessary	60.9	Average	22.7	Unnecessary	3.6	Highly unnecessary	0.0
Method of data sharing between evaluation programs	Evaluator	Convert in DB	44.4	Unification of formats	34.6	Disclosure of data	17.3	Shared among counselors	3.7	-	-
Degree of differences in contents of evaluation data among evaluation systems	Subject	Very different	0.9	Different	28.4	Average	45.4	Similar	24.4	Very similar	0.7
Degree of differences in formats of evaluation data among evaluation systems	Subject	Very different	2.6	Different	41.2	Average	42.6	Similar	15.8	Very similar	0.4

order applies hereinafter). It can be resumed that the evaluator feels burdened with reviewing a heavy data load and the subject feels additional pressures since it is conducted regardless of the task evaluation.

On the question asking the level of satisfaction over the evaluation result, more respondents were generally satisfied (32.7%, 42.6%) than unsatisfied (24.5%, 15.1%) and the subjects showed a somewhat higher level of satisfaction than the evaluators. It is judged that the higher level of satisfaction among subjects over evaluators is very unique where the subjects recognized that they achieved satisfactory result; whereas, the evaluators judge that the overall evaluation was higher than what they expected. This could be resulted from overstated markings due to the solicitous judgment of some evaluators.

The result of inquiry into the experience of repeated evaluations on more than two National R&D Program evaluations at the same time showed that 31.9% had the experience and that they occurred most frequently in the 4th quarter (46.2%). This corresponds to the general speculation that the evaluation would be concentrated in particular time.

About the appropriateness of evaluation period, more people answered that it was appropriate, but 24 evaluators and 65 subjects answered 'inappropriate.' The evaluators chose 2nd quarter (45.8%) and 1st quarter (29.2%); while the subject answered 1st quarter and 3rd quarter (both 33.8%). It appears that they chose to avoid the 4th quarter since it overlaps with the conclusion period of programs and tasks and many evaluation programs such as HR evaluation are all concentrated in that period.

To the question on repetition among evaluation programs, the respondents thought that task evaluation and program evaluation is highly repetitive (evaluator: 44.5%; subject: 43.3%). This appears to be the result of their experience of submitting identical data on both task evaluation and program evaluation. It can be deduced that the level of difference in contents of evaluation submitted to each evaluation system and the answer 'different' was lower (29.3%); whereas, the level of difference in the format showed that the answer 'different' was higher (41.2%). That is, the subject experienced difficulty in writing similar contents in a different format.

Next, in terms of evaluators, many felt that the program evaluations and institution evaluations were redundant (22.7%). It seems that, since there are comment elements in partial indices in program evaluation and institution evaluation, they would have identified some repetition. The question on the necessity of sharing evaluation data among evaluation systems to the evaluators proved the need to share (73.6%)

3.3.2 Recognition on Pressures of the National R&D Program Evaluation (Table 7)

On the question whether the evaluation task pressures the research task, 41.8% of the evaluators and 47.4% of the subject responded positive, which shows that it is burdensome for both parties. Moreover, the subjects felt more pressured than the evaluators, illustrating that subjects feel more pressured. Also, among those who answered the evaluation task on being pressured (46 evaluators, 270 subjects), both parties chose various evaluation data (78.3%, 57.8%) as the most pressuring element followed by writing the report (19.6%, 34.1%), revealing that writing evaluation data was very burdensome.

As for time consumed in conducting evaluation, the evaluators answered 'less than 3 days' most frequently (26.4%), followed by10-19 days (20.0%) and 5-9 days (19.1%) for an average of 15.1 days. Subjects

Table 7 Recognition on Pressures of National R&D Program Evaluation

Questions	Subject	Responses											
		Item1		Item 2		Item 3		Item 4		Item 5		Item6	
Level of evaluation task pressure on the research and the cause of pressure	Evaluator (N=46)	Highly burdensome	4.5	Burdensome	37.3	Average	40.0	Not burdensome	16.4	No burden at all	1.8	-	-
		Evaluation data	78.3	Report	19.6	Meeting	0.0	Business Trip	0.0	Others	2.2	-	-
	Subject (N=270)	Highly burdensome	51.0	Burdensome	42.3	Average	36.7	Not burdensome	14.7	No burden at all	1.2	-	-
		Evaluation data	57.8	Report	34.1	Meeting	4.4	Business Trip	3.3	Others	0.4	-	-
Time consumed in evaluation and evaluation data	Evaluator	Less than 3 days	26.4	3-4 days	10.9	5-9 days	19.1	10-19 days	20.0	20-49 days	16.4	More than 50 days	7.3
	Subject	Less than 3 days	20.9	3-4 days	25.3	5-9 days	16.3	10-19 days	13.0	20-49 days	10.4	More than 50 days	14.2

answered 3-4 days the most (25.3%), followed by less than 3 days (20.9%), and 5-9 days (16.3%) for 12.5 days on average. Considering that the number of days that actually are available for research annually is around 250 days, some 5% of research days are spent in evaluation.

3.3.3 Recognition of Reflecting Research Characteristics on the Evaluation (Table 8)

For the question regarding the level of evaluation method that reflects the characteristics of research activity, many evaluators (40.9%) felt it is reflected and subjects chose 'average' (43.7%) the most. Among the respondents who answered that it did not reflect that characteristics (21 evaluators, 128 subjects), they chose discerning technological characteristics (38.1%, 44.5%) and the characteristics in R&D phase (33.3%, 40.6%).

Next, for the question whether it is desirable to voluntarily provide performance goals and indices and verify it, the answer "it is positive (55.4%, 63.99)," greatly outweighed the negative (9.1%, 8.7%) illustrating that both parties generally found desirable voluntary provision and preliminary verification system on the performance plans.

For the question as to whether the performance goals and index reflect objectives and characteristics of the activity, the opinion that it is reflected (40.0%, 51.2%) was higher and subjects found it more positive than the evaluators. Among the respondents who answered that it does not reflect the objectives and characteristics (16 evaluators, 59 subjects), many people chose "inappropriate method of verification" (50.0%, 40.7%) followed by lack of expertise of the verifier (31.3%, 37.3%).

3.3.4 Recognition Regarding Utilization of Evaluation Results (Table 9)

The level of researcher's use of evaluation results among the evaluators were in the order of 'average' (50.0%), 'used' (26.3%) and 'not used' (23.7%) revealing that the majority used the result but the

Table 8 Recognition of Reflecting Research Characteristics on the Evaluation

Questions	Subject	Responses									
		Item1		Item 2		Item 3		Item 4		Item 5	
Level of evaluation method reflecting R&D activity characteristics and the cause of failure in reflecting	Evaluator (N=21)	Well reflected	0.9	Reflected	40.0	Average	40.0	Not reflected	15.5	Not reflected at all	3.6
		Technological characteristics	38.1	Development stage	33.3	Industrial characteristics	19.0	Regional characteristics	9.5	Others	0.0
	Subject (N=128)	Well reflected	1.2	Reflected	32.6	Average	43.7	Not reflected	20.5	Not reflected at all	1.9
		Technological characteristics	44.5	Development stage	40.6	Industrial characteristics	9.4	Regional characteristics	3.9	Others	1.6
Current performance plan verification system	Evaluator	Highly desirable	3.6	Desirable	51.8	Average	35.5	Inappropriate	9.1	Highly inappropriate	0.0
	Subject	Highly desirable	6.5	Desirable	51.8	Average	27.4	Inappropriate	8.2	Highly inappropriate	0.5
Level of performance goals and index reflecting the characteristics of R&D activity and the cause of failure in reflecting	Evaluator (N=16)	Well reflected	1.8	Reflected	38.2	Average	45.5	Not reflected	14.5	Not reflected all	0.0
		Verification method	50.0	Verifier's Expertise	31.3	Verification period	6.3	External environment	0.0	Others	12.5
	Subject (N=59)	Well reflected	3.7	Reflected	47.5	Average	38.4	Not reflected	9.6	Not reflected all	0.7
		Verification method	40.7	Verifier's Expertise	37.3	Verification period	11.9	External environment	5.1	Others	5.1

Table 9 Recognition Regarding Utilization of the Evaluation Result

Questions	Subject	Responses									
		Item1		Item 2		Item 3		Item 4		Item 5	
Level of researcher's utilization of evaluation result	Evaluator	Highly used	4.5	Used	21.8	Average	50.0	Not used	18.2	Not used at all	5.5
Reliability of evaluation result	Evaluator	Very high	2.7	High	26.4	Average	52.7	Low	13.6	Very Low	4.5
Fairness of evaluation result	Evaluator	Very high	4.5	High	28.2	Average	46.4	Low	16.4	Very Low	4.5
Level of internal use and utilization details of evaluation results	Subject	Highly used	4.7	Used	44.0	Average	34.0	Not used	19.6	Not used at all	2.3
	(N=251)	Program planning	65.3	Individual evaluation	21.9	Incentive	10.8	Following year's salary	1.2	Others	0.8
Opinion on associating evaluation results to the following year's budget	Evaluator	Highly agreed	6.4	Agreed	40.9	Average	33.6	Not agreed	19.1	Not agreed at all	0.0
	Subject	Highly agreed	6.8	Agreed	51.4	Average	30.2	Not agreed	10.7	Not agreed at all	0.9
Experiences of different results according to the host of evaluation and the cause	Subject	Positive	51.2	Negative	15.1	N/A	33.7	-	-	-	-
	(N=292)	Evaluation counselor	71.6	Evaluation method	18.2	Performance planning	8.9	Others	1.0	-	-
Experiences of different results despite similarities with previous year's data and the cause	Subject	Positive	31.4	Negative	28.8	N/A	39.8	-	-	-	-
	(N=179)	Consistency	82.1	Other conditions	11.7	Performance planning	6.1	-	-	-	-
Experiences of different results according to the host of evaluation and the cause	Evaluator	Positive	55.5	Negative	24.5	N/A	N/A	-	-	-	-
	(N=61)	Evaluation counselor	77.0	Performance planning	14.8	Other conditions	6.6	Others	1.6	-	-

percentage that did not using was still high.

The question on the reliability of the evaluation result for the evaluators showed the following order: 'average' (50.0%); 'reliable' (29.1%), and 'not reliable' (18.1%). In regards to fairness, results were in the order of average (46.4%), fair (32.7%) and not fair (20.9%),illustrating the general recognition that there is no problem in reliability and fairness.

Results for the question on the level of internal use of the evaluation result were in the following order: used (44.0%), average (34.0%) and not used (21.9%). The 251 respondents who used the result responded in the order of program planning (65.3%), individual evaluation (21.9%), and incentive (10.8%).

For opinions on associating the evaluation result on the budget, a majority of the respondents (47.3%, 58.2%) agreed on the idea, with average (33.6%, 30.2%), and negative opinion (19.1%, 11.6) following next.

Results of responses to experiences of different evaluation results according to the host of evaluation showed that 51.2% of the subjects had the experience and 71.6% chose lack of expertise and fairness of the evaluation counselor as the cause, revealing that reliability of the evaluation result was low. Similarly, about the experience of achieving different results

even though it is similar to the evaluation data of the previous year, many respondents (55.5%, 31.4%) had the experience and chose lack of consistency (77.0%, 82.1%) as a cause, which also shows low level of reliability.

When a question about the experiences of different results according to the evaluation host was asked to the evaluator, 55.5% answered positive and also pointed out a lack of consistency.

3.3.5 Recognition of the Evaluation Council (Table 10)

On the question "which is more important between the expertise and fairness of the evaluation counselors," the answer, "expertise is more important" (50.0%, 48.2%) was highest, "fairness is more important" (37.3%, 39.6%) came next, and "the same" (12.7%, 12.1%) trailed behind. Both parties found expertise more important than fairness.

Regarding the expertise of the evaluation counselor, 'average' (47.3%, 42.5%) was the highest, 'high level of expertise' was next (34.5%, 38.2%), and 'low level of expertise' (18.2%, 19.3%) came last. Also, the respondents who chose low level of expertise were asked to identify the cause. Evaluators responded as follows: lack of experience (45.0%); lack of understanding (20.0%) and standards of choosing professionals (15.0%). For the subjects, the response, "lack of experience in related fields and standard of choosing professionals" was the highest at 36.4%, non-specialized area was next (16.4%), and lack of understanding the data (8.2%) followed next, showing difference in perspectives among personnel beyond mere lack of experience.

As for fairness of evaluation counselor, the answer 'average' (57.3%, 43.5%) was the highest; 'high' (30. 0%, 43.5%), and 'low' (12.7%, 20.5%) followed next, which illustrates the judgment that there is no significant problems in fairness. Also, among the respondents (14 evaluators, 117 subjects) who chose 'low in fairness,' the majority of respondents chose 'absence of systems' (50.0%), and 'lack of qualification' (both 29.9%), school ties (29.1%), and regionalism (5.1%), revealing showing differences in perspectives.

3.3.6 Improving the Performance Evaluation System

3.3.6.1 Recognition of Improving the Evaluation System (Table 11)

Table 10 Recognition of the Evaluation Counsel

Questions	Subject	Responses									
		Item1		Item 2		Item 3		Item 4		Item 5	
Important element between 'expertise' and 'fairness' of the evaluation counselor	Evaluator	Expertise	50.0	Fairness	37.3	-	-	-	-	-	-
	Subject	Expertise	48.2	Fairness	39.6	-	-	-	-	-	-
Level of expertise of evaluation counselor and cause of low level of expertise	Evaluator	Very high	3.6	High	30.9	Average	47.3	Low	17.3	Very low	0.9
	(N=20)	Lack of experience	45.0	Standard of choice	15.0	Non-specialized	10.0	Lack of understanding	20.0	Others	10.0
	Subject	Very high	4.0	High	34.2	Average	42.5	Low	16.5	Very low	2.8
	(N=110)	Lack of experience	36.4	Standard of choice	36.4	Non-specialized	16.4	Lack of understanding	8.2	Others	2.7
Level of fairness of evaluation counselor and cause of low fairness	Evaluator	Very high	3.6	High	26.4	Average	57.3	Low	11.8	Very low	0.9
	(N=14)	Qualification	14.3	Absence of systems	50.0	School ties	14.3	Regionalism	7.1	Others	14.3
	Subject	Very high	4.2	High	31.8	Average	43.5	Low	17.5	Very low	3.0
	(N=117)	Qualification	29.9	Absence of systems	29.9	School ties	29.1	Regionalism	5.1	Others	6.0

Table 11 Recognition on Improving the Evaluation System

Questions	Subject	Responses									
		Item1		Item 2		Item 3		Item 4		Item 5	
Most important element regarding improving the evaluation system (multiple choice)	Evaluator	Alleviating the pressures on materials	97.3	Merging similar evaluations	70.9	Reducing the number of evaluations	83.6	Shortening of evaluation periods	41.8	Others	6.4
	Subject	Alleviating the pressures on materials	89.1	Merging similar evaluations	81.8	Reducing the number of evaluations	76.7	Shortening of evaluation periods	45.4	Others	6.3
Level of helpfulness of triennial evaluation system for alleviating evaluation pressures	Evaluator	Very helpful	29.1	Helpful	45.5	Average	18.2	Not helpful	4.5	Not helpful at all	2.7
	Subject	Very helpful	27.0	Helpful	51.9	Average	13.7	Not helpful	6.7	Not helpful at all	0.7

As for improvement of evaluation systems, respondents regarded reduction of submission burdens (97.3%, 89.1%) most importantly, and evaluators responded in the order of reduction of the number of evaluations (83.6%) and merging similar evaluation programs (70.9%); whereas the subject chose merging similar evaluation programs (81.8%), reduction of the number of evaluation (76.7%), and shortening of period (45.4%) which showed that they found alleviating the pressures in submitting data was important.

On the question whether to revise the evaluation cycle from one year to three years, the answer that it is helpful (74.6% of evaluators, 78.9% of subjects) was highest.

3.3.6-2 Evaluator Inquiries on Improving Evaluation Systems (Table 12)

The result of investigating opinions on improving current evaluation systems among the evaluators indicated that the voices that require 'evaluations according to program characteristics' (15.4%) was highest. Other answers included 'diversification of evaluation methods' (9.1%), 'requires concrete evaluations' (3.6%), and 'index development through grouping' (2.7%). In sum, rather than have a unified system of evaluation with identical standards, it is better to diversify the evaluation methods by taking diverse environments, i.e. type, process, period and size, into consideration since the subject of evaluation has diverse characteristics according to program, task and institution. Meanwhile, merging needs to enable the absorption of unique characteristics of individual evaluation subjects with a conception that the evaluation by its program characteristics is a trade-off to the merging of evaluation systems or utilization of comprehensive data.

Moreover, responses related to alleviating evaluation pressures totalled 15.4%, including merging of identical/similar evaluations (7.3%), simplification of evaluations (4.5%), and improvement of evaluation cycles (3.6%). In contrast, a minority opinion (2.7%) was concerned about the side effects of triennial evaluations that could actually aggrevate the burden.

Next, 14.5% of the respondents demanded assurance of expertise during evaluation, indicating a demand for reinforcing the reliability of evaluation results. On

Table 12 Evaluator Inquiries on Improving Evaluation Systems (For Evaluators)

Majority Responses (%)			
Evaluations according to program characteristics	15.4	Reinforcement of consulting function	3.6
Alleviation of evaluation pressures	15.4	Insurance of evaluation fairness	2.7
Insurance of evaluation expertise	14.5	Providing benefits by evaluation results	2.7
Merging of identical/similar evaluation systems	7.3	Concerns on triennial evaluations	2.7

the other hand, only 2.7% sought greater fairness, indicating a higher overall demand for expertise over fairness.

3.3.6.3 Evaluation Subject Inquiries on Improving the Evaluation System (Table 13)

The results from an examination of the opinions of 570 subjects showed substantial differences from those of the evaluators. First, the respondents who demanded 'ensuring expertise of the evaluation' were highest (19.3%). Insurance of evaluation fairness followed next (18.4%), showing the desperate need for improvement in both expertise and fairness. The demand for enhancement of expertise was higher than the need for fairness, a matter that is directly associated with the reliability of evaluation results. This could stem from the uncertainty regarding the level of expertise among evaluators due to the characteristics of the research institution recognized as a group of experts in the specific research area.

The answers – simplification of evaluations (10.0%), improvement of evaluation cycle (3.2%), and merging of identical/similar systems (2.5%) – followed, showing concerns on evaluation burdens. Then there were the demands for satisfying the diversity of evaluation subjects– diversification of evaluation methods (6.0%); and detailed evaluation (3.2%). This is different from the responses of the evaluators, who presented higher demands for diversification of evaluation methods. It appears that the subjects found the reduction of evaluation pressure to be more urgent Other opinions included satisfaction with the current evaluation system (4.0%). The cause of this response could have derived from the fact that the evaluation pressure is actually aggravated due to the lack of consistency and change of detailed directions, although improvement evaluation systems is taking place annually.

4. Implications for Developmental Settlement of Performance-Based Evaluation Systems

We have looked at the current status of evaluation systems through analysis and reference reviews of National R&D Program evaluation systems, composed issues from experiences as a questionnaire, and investigated and analyzed the status of recognition of evaluators and evaluation subjects. Through this process, it was possible to grasp the level of recognition of concerned parties as well as the current status of evaluation systems on National R&D Programs. This chapter is intended to derive political implications and methods of improvement to be referenced in future system revisions through comprehensive analysis of questionnaire results.

4.1 Political Implications on National R&D Program Evaluation Systems

4.1.1 Necessity of Improving Self and Meta Evaluation Systems

In contrast to the evaluators, 60.5% of the subjects imparted more importance on task evaluations than program evaluations. On the other hand, both parties chose program evaluations as the most pressured evaluation system. Subsequently, the subjects recognized the self and meta evaluation systems as highly pressured with a low level of importance.

This might have caused by the problem of where the responsibility lies since the result of program evaluation is directly considered as the evaluations

Table 13 Evaluation Subject Inquires on Improving the Evaluation System (For evaluation subjects)

Majority Responses (%)			
Assurance of evaluation expertise	19.3	Satisfied with current evaluation systems	4.0
Assurance of evaluation fairness	18.4	Unification of evaluation standards	1.9
Alleviation of evaluation pressures	15.7	Selection of evaluation counselor through establishment of a evaluation counselor pool	2.7
Evaluations according to the program characteristics	9.2	Providing incentives according to evaluation results	1.1

on personnel in charge of the department; whereas, the task evaluation directly affects the personnel in charge of the research. Moreover, the fact that the result of evaluation on the program is biased toward performance evaluation and fails to provide necessary information to the parties that conduct particular research tasks could be another factor.

However, it is necessary to enhance the level of recognition on the program evaluation through active promotion in the future. This is because the program is a R&D management unit on the national level, and at the same time, a means of policy execution and the feedback on evaluation results has significant impact on the program unit, not to mention the task unit

Furthermore, there are cases of repeated submission of data since the meta evaluation takes place as a format to re-review the self-evaluated programs. This is because the meta evaluation is conduced as a form to revaluating self-evaluation results. Thus, it is necessary to devise a plan to improve systems for more effective evaluations since this could bring disputes regarding redundancy in evaluation.

4.1.2 Need for preparing plans to alleviate pressures in writing evaluation materials

A large number of personnel who felt pressured answered that the recording of evaluation data and writing the report were the most burdensome. Moreover, the average time consumed in writing the evaluation was 15.1 days for the evaluators and 12.5 days for the subjects, showing that a large amount of time is spent in writing evaluations. By simple calculation, it means that if a person were evaluated twice a year, the person would consume approximately one month per year in writing the evaluation, which could impede research. Moreover, the subjects felt that the similarity in content but disparity in formats added to of the perception of redundancy.

Although written evaluations are unavoidable, avoiding redundancy is desirable. For this, determining a similar constituent among evaluation systems through analysis and investigation is required. If at all possible, developing standardized formats for similar data and minimizing repeated writing through shared data should be undertaken.

Dissertations, patents and technology transfers are representative of repeated and similar data. As of now, the performance information created by each research host is managed comprehensively and shared through NTIS. There lie the limitations in which the subject of sharing is limited to only some performances and it is utilized only in self and meta evaluations without any particular regulators. Therefore, expanding scope of sharing in the future and preparing related regulations is necessary.

4.1.3 Necessity of establishing evaluation systems for each program type to reinforce expertise in evaluation systems

The majority of the respondents answered that evaluation systems reflect research characteristics; whereas, 19.1% of evaluators and 22.4% of the subjects still feel they are insufficient. Moreover, both the evaluators and subject found expertise is more important than the fairness in evaluation. Considering this result, reinforcing evaluation expertise to reflect characteristics of the program in current evaluation systems is essential.

As the evaluation system is improved in the direction of reinforcing qualitative evaluations, more customized and consulting type evaluations are being carried out. For this, the portion of quantitative evaluations has increased recently; whereas, the qualitative evaluation elements which enables customized evaluation is decreasing. However, to reinforce professionalism in the program evaluation, development and introduction of qualitative evaluation elements is required to integrate the quantitative and qualitative evaluations of the experts.

Currently, the evaluation result is derived by calculating the relative ranking to the overall programs. This underlies the risk of inhibiting the individuality of each program. Thus, systematic improvement must take place to complement the relative evaluation systems of today. For instance, categorizing similar programs and drawing relative rankings within those rather than comparing them with entire programs can be considered. There is a good example which similar

programs are grouped into one program group in in-depth evaluations.

On the other hand, in-depth evaluations are conducted by developing specialized performance creation logical models, and comprehending and analyzing the performance in depth. Doing so could contribute to enhancing evaluation professionalism that does not discern program characteristics. Nonetheless, since in-depth evaluation involves enormous time and cost as well ashuman and research resources, it to putting more effort in expanding those gradually is necessary.

4.1.4 Enhancement of Evaluation Results Utilization

Although the majority of the respondents utilized the evaluation result internally, more than 20% of the respondents answered that they do not utilize the results. Considering that the current 'Law on Performance Evaluation' stipulates arbitration and allotment of program budgets, amendment and supplementation of program promotion plans, improvement of researcher's employment conditions and research environment, and rewards on outstanding outcomes, the level of result utilization is still very low.

When divided into internal and external use, in terms of external use, it is used effectively by the government where budgets are concerned such as increasing or decreasing the budgets according to the results of the self and meta evaluation; while the use inside each department and project groups largely stays at mere reformation of criticized elements, apart from active utilization concerning execution of program restructuring or coordination of research portfolio.

One of the main reasons behind the low level of utilization is that the focus of evaluation is concentrated on reflecting the results on budgets. As the Ministry of Strategy and Finance, the primary department authorizing national finances, executed the role of evaluation colligation, budgetary usage has been amalgamated but other parts has been stagnated or has not shown marked advances.

However, it is not pertinent to associate evaluation results directly with budgetary considerations in the case of programs with national gravity – development of original technology that requires long-term investment or green technology – and programs that should be perceived from comprehensive science technology – consolidation programs or next generation food creation programs. Rather than simply associating evaluation results with budgetary decisions, it is important to improve program structures or operations according to criticism or recommendations. Thus an effort should be taken by the department to submit particular execution planning according to the evaluation result and the colligating department to establish a monitoring and supervising system by establishing a system that coerces program restructuring according to evaluation results.

4.2 Differences in Recognition of Evaluators and Evaluation Subject (Table 14)

In this study, the questionnaire sheet was composed and questioned each party separately with the assumption that there would be difference in recognition between the parties as well as the current status of overall recognition. As a result, the respondents illustrated different opinions on 7 items from a total of 37 items (18.9%); however, generally, they appear to be similar.

The summary of areas of difference in recognition is as follows. First, for the question on the important system, subjects imposed more importance on task evaluation compared to the evaluators. For the question on the appropriate period of evaluation, evaluators chose the 2nd quarter, while the subject chose the 1st and 3rd quarters. Also, the time consumed in writing evaluation materials was longer by 2.6 days on average for evaluators compared to subjects. On the question whether the method of evaluation reflects program characteristics, the evaluator responded positively, while the subjects responded with a response of average. For the cause of low level of expertise among evaluation counselors, the evaluators chose lack of experience; whereas, the subjects chose lack of experience and problems with the selection standard. For the reason of low level of fairness, the evaluator identified problems with the system, while the subject indicated problems

Table 14 Differences in Recognition of Evaluators and Evaluation Subject

Item	Evaluator	Evaluation Subject	Remarks
• Important evaluation	Task/self evaluation	Task evaluation	-
• Appropriate evaluation period	2nd quarter	1st and 3rd quarter	-
• Time consumed in preparing evaluation materials	15.1 days	12.5 days	-
• The level of reflection of program characteristics in the evaluation method	Reflected	Average	-
• Cause of low level of expertise of evaluation counselors	Lack of experience	Lack of experience and selection standard	-
• Cause of low level of fairness of evaluation counselors	Problems with evaluation systems	Evaluation systems and quality of evaluation counselors	-
• Urgent improvement request particulars	Specialization, alleviation of pressure, expertise	Expertise, fairness, alleviation of pressure	Short answer

with the system and qualifications which showed difference in their selection.

Lastly, for the short answer question on contents that requires urgent reformation of systems, the evaluator chose reflection of program characteristics, alleviation of evaluation pressures, and enhancement of expertise; whereas, the subject selected enhancement of evaluation expertise and fairness, and alleviation of evaluation pressures presenting difference in recognition.

The results of this study could be used as basic information to reference and utilize systemic improvement of National R&D Programs to settle performance-based evaluation systems in R&D program evaluation. It is expected that there will be a high value of utilization as basic information in understanding how the recognition of concerned parties are changing and in what direction it is flowing according to the reformation of systems when the transition process of recognition is investigated and analyzed regularly in the future.

Table 15 Main Reponses from the Questionnaire Result and Result of Statistic Significance Analysis

Class	Question	Subject	Main Responses (%)		Difference in Recognition	Statistics
R&D activities and government subsidized (RI) evaluation status	The most important evaluation	Evaluator	Task evaluation, Program evaluation	47.3, 44.5	None	χ^2=7.525*
		Subject	Task evaluation	60.5		
	The most burdensome evaluation	Evaluator	Program evaluation	60.9	None	χ^2=78.064***
		Subject	Program evaluation	46.5		
	Level of satisfaction on evaluation results	Evaluator	Satisfied	32.7	None	F=5.913*
		Subject	Satisfied	42.6		
	Experience of being subjected to multiple evaluations at the similar period	Subject (N=182)	Yes	31.9	-	-
			4th Quarter	46.2	-	-
	Appropriateness of current evaluation periods and the time	Evaluator	Inappropriate	21.8	None	F=3.52(n.s)
		Subject	Inappropriate	11.4		
		Evaluator(N=24)	2nd Quarter	45.8	Quarterly difference	χ^2=8.934*
		Subject(N=65)	1st ,3rd Quarter	33.8 each		

Table 15 Main Reponses from the Questionnaire Result and Result of Statistic Significance Analysis (cont'd)

Class	Question	Subject	Main Responses (%)		Difference in Recognition	Statistics
R&D activities and government subsidized (RI) evaluation status	The evaluation system which the evaluation content is thought to be repeated	Evaluator	Task evaluation and Program evaluation	44.5	None	χ^2=23.244***
		Subject	Task evaluation and Program evaluation	43.3		
	Necessity of data sharing between evaluation programs	Evaluator	Necessary	73.6	-	-
	Method of data sharing between evaluation programs	Evaluator	Converting into a DB	44.4	-	-
	The degree of differences in contents of evaluation data among evaluation systems	Subject	Average	45.4	-	-
	The degree of differences in formats of evaluation data among evaluation systems	Subject	Average	42.6	-	-
Evaluation Pressure Factor	The degree of pressures of evaluation tasks on research tasks	Evaluator	Pressured	41.8	None	F=1.03(n.s)
		Subject	Pressured	47.4		
	The most pressured factors in conducting evaluations	Evaluator(N=46)	Various evaluation material	78.3	None	χ^2=10.900*
		Subject(N=270)	Various evaluation material	57.8		
	The time spent in writing evaluation and data	Evaluator	15.1 days (avg.)	100.0	3 days	-
	The time spent in writing evaluation and data	Subject	12.5days (avg.)	100.0		
Reflection of Research Characteristics	Level of evaluation methods in reflecting characteristics of R&D activities	Evaluator	Reflected	40.9	Reflected	F=0.988 (n.s)
		Subject	Average	43.7		
	Cause of failure of evaluation methods in reflecting characteristics of R&D activities	Evaluator(N=21)	R&D characteristics of each technology	38.1	None	-
		Subject(N=128)	R&D characteristics of each technology	44.5		
	Current performance plan verification system	Evaluator	Desirable	55.4	None	F=2.026 (n.s)
		Subject	Desirable	63.9		
	Level of performance goals and index reflecting the characteristics of R&D activity	Evaluator	Reflected	40.0	None	F=4.597 *
		Subject	Reflected	51.2		
	The cause of failure in reflecting the characteristics of R&D activity	Evaluator(N=16)	Inappropriate verification method	50.0	None	-
		Subject(N=59)	Inappropriate verification method	40.7		

Table 15 Main Reponses from the Questionnaire Result and Result of Statistic Significance Analysis (cont'd)

Class	Question	Subject	Main Responses (%)		Difference in Recognition	Statistics
Utilization of evaluation results	Level of researchers' utilization of evaluation results	Evaluator	Average	50.0	-	-
	Reliability of evaluation results	Evaluator	Average	52.7	-	-
	Fairness of evaluation results	Evaluator	Average	46.4	-	-
	Level of internal utilization of evaluation results	Subject	Used	44.0	-	-
	Contents of internal utilization of evaluation results	Subject(N=251)	Reflected on program plans	65.3	-	-
	Opinions on integrating evaluation results onto the budget of the following year	Evaluator	Positive	47.3	None	F=4.514 *
	Opinions on integrating evaluation results onto the budget of the following year	Subject	Positive	58.2		
	Experiences and causes on different evaluations results from different evaluation authorities	Subject	Yes	51.2	-	-
		Subject(N=292)	Problem of evaluation counselor	71.6		
	Experiences and causes on different results on the data similar to the previous year	Subject	Yes	31.4	-	-
		Subject(N=179)	Lack of consistency	82.1		
	Experiences and causes on different evaluations results from different evaluation authorities	Evaluator	Yes	55.5	-	-
		Evaluator(N=61)	Lack of consistency	77.0		
Evaluation Counsel	Important element between 'expertise' and 'fairness' of the evaluation counselor	Evaluator	Expertise	50.0	None	-
		Subject	Expertise	48.2		
	Expertise of evaluation counselors	Evaluator	average	47.3	None	F=0.015 (n.s)
		Subject	average	42.5		
	Cause of low level of expertise of evaluation counselors	Evaluator(N=20)	Lack of experience	45.0	Expert Selection	-
		Subject(N=110)	Lack of experience, expert selection standard	36.4 each		-
	Fairness of evaluation counselors	Evaluator	average	57.3	None	F=0.143 (n.s)
		Subject	average	43.5		
	Cause of low level of fairness of evaluation counselors	Evaluator (N=14)	Evaluation system	50.0	Qualification	-
		Subject(N=117)	Evaluation system, qualification	29.9 each		
Improving the Evaluation System	Most important element in regards to improving the evaluation system (multiple choice)	Evaluator	Alleviating the pressures on materials	97.3	None	-
		Subject	Alleviating the pressures on materials	89.1		
	Level of helpfulness of triennial evaluation system on alleviating evaluation pressures	Evaluator	Helpful	74.6	None	-
		Subject	Helpful	78.9		

n.s=non-significance, * p<.05, ** p<.01, *** p<.0.001, N : Evaluator - 110, Subject - 570

4. Reference

1. National Science and Technology Council (NSTC) (1999), "Results of research, analysis and evaluation of Government R&D Program"
2. Ministry of Education, Science & Technology (MEST) (2005), "Research on examination, analysis and evaluation support for Government R&D program in 1999"
3. Ministry of Education, Science & Technology (MEST) (2005), "Law on evaluation and management of performances in Government R&D Program"
4. Seol, Sang-Su (2005), "National science and technology evaluation system establishment planning", Korea Institute of Science & Technology Evaluation and Planning (KISTEP)
5. Yang, Hee-Seung (2002), "Status of National R&D Programs (2002)", Korea Institute of Science & Technology Evaluation and Planning (KISTEP)
6. Oh, Dong Hoon; Son, Byeong Ho, et al (2007), "Study on compilation of textbook on theories and practical aspects of evaluation of R&D Program", Korea Institute of Science & Technology Evaluation and Planning (KISTEP)
7. Oh, Dong Hoon (2006), "Study on establishment of National R&D Evaluation System", Korea Institute of Science & Technology Evaluation and Planning (KISTEP)
8. Lee, Ki-Jong; Hwang, Deok-Gyu; Kim, Seong-Jin; Kim, Seung-Tai (2008), "Exploratory study on enhancement of efficiency of evaluation system for R&D performances – with focus on examination of recognition of researchers on evaluation system", 2008 Autumn Seminar of Korea Technology Innovation Society
9. Lee, Sang-Yeop; Jeong, Geun-Ha (2007), "Study on deduction and analysis of issues for efficient evaluation of Government R&D Program", Korea Institute of Science & Technology Evaluation and Planning (KISTEP)
10. Lee, Chan-Gu (2009), "Empirical analysis of evaluation of research institutes: with focus on the differences in the recognition of the participants in evaluation", Journal of Korea Technology Innovation Society, Volume 12, No. 1, pp.36~69
11. Lee, Chan-Gu (2008), "Study on the analysis of differences in and integrated goals of evaluation of government subsidized research institutions"
12. Lee, Hyung-Hu (2005), "Analysis on evaluation systems on National R&D Programs", Korean Association for Policy Studies, pp. 541-565
13. Korea Institute of Science & Technology Evaluation and Planning (KISTEP). (2001), "Investigation, analysis, and evaluation result of 2001 National R&D Programs", National Science and Technology Council (NSTC)
14. Korea Institute of Science & Technology Evaluation and Planning (KISTEP). (2004), "Investigation, analysis, and preliminary arbitration result of 2004 National R&D Programs", National Science and Technology Council (NSTC)
15. Korea Institute of Science & Technology Evaluation and Planning (KISTEP). (2005), "Investigation, analysis, and evaluation result of 2005 National R&D Programs: Evaluation" National Science and Technology Council (NSTC)
16. Korea Institute of Science & Technology Evaluation and Planning (KISTEP). (2006), "Basic plan for evaluation of R&D performances, and basic plan for management and utilization of research performances", National Science and Technology Council (NSTC)
17. Korea Institute of Science & Technology Evaluation and Planning (KISTEP). (2008), "Basic Science & Technology Basic Plan towards becoming a global leader - 577 Initiative", Ministry of Strategy & Finance (MOSF), Ministry of Education, Science & Technology (MEST) and Ministry of Education & Science Technology, et al
18. Korea Institute of Science & Technology Evaluation and Planning (KISTEP). (2009), "2009 Comprehensive guideline for National R&D Programs"
19. Korea Institute of Science & Technology Evaluation and Planning (KISTEP). (2009), "2009 National R&D Program Meta Evaluation Plans", Ministry of Strategy & Finance (MOSF)
20. Korea Institute of Science & Technology Evaluation and Planning (KISTEP). (2009), "2009 Performance Evaluation Execution Plan", Korea Institute of Science & Technology Evaluation and Planning (KISTEP).
21. Korea Institute of Science & Technology Evaluation and Planning (KISTEP). (2009), "2009 National R&D Program Self Evaluation Manual", Ministry of Strategy & Finance (MOSF)
22. National Technical Information Service (NTIS): http://www.ntis.go.kr