기술혁신학회지 제17권 4호
2014년 12월 pp.703~732

The Design and Application of a Meta-evaluation Model for National R&D Programs†

국가연구개발사업 메타평가모형 설계와 응용

류영수(Youngsoo RYU)*, 최병대(Byungdae CHOI)**, 최상옥(Sang Ok CHOI)***

목 차

국 문 요 약

국가연구개발사업에 대한 메타평가는 과학기술수요와 목표와 연계된 국가전체의 연구개발시스템의 틀 안에서 구축되어야 한다. 연구개발사업평가시스템은 투입, 수행, 산출, 활용의 전 과정으로 이해될 수 있다. 전문가 24명이 참여한 전문가 델파이기법을 통하여 24개의 평가지표가 개발되었다. 메타평가 모델은 중요한 관점 8개의 하위시스템을 포함하고 있다. 8개 하위시스템에는 기획의 적실성, 정보의 충실성, 평가투입에서 평가자의 적정성, 평가수행과정과 방법의 적절성, 평가활용보고와 응용의 유용성 등이 포함된다. 메타평가모형의 활용방법으로서 평정점수의 평균은 5점척도 기준으로 2.73이며, 이는 평가점수의 중앙값 3.0보다 낮게 나타났다. 평가결과의 환류는 연구개발프로그램의 개선에 특별히 강화될 것을 제언한다.

핵심어 : R&D 프로그램 평가, 메타평가, 평가 척도, 정부역량

※ 논문접수일: 2014.9.8, 1차수정일: 2014.11.24, 게재확정일: 2014.12.1
* 한국과학기술기획평가원 선임연구위원, soory@kistep.re.kr, 02-589-2838
** 한양대학교 행정학과 교수, choibd@hanyang.ac.kr, 02-2220-0831
*** 고려대학교 행정학과 부교수, sangchoi@korea.ac.kr, 02-3290-2284, 교신저자
† This study was partially supported by a Korea University Grant and the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2012-2012S1A3A2033282).

ABSTRACT

Meta-evaluation for national R&D programs should be established within the framework of whole R&D system linked with science and technology needs and objectives. The R&D program evaluation system can be understood as a circulation of input, implementation, output, and utilization. In the result of the Delphi survey, which consisted of twenty-four experts (performed in three rounds), twenty-four indices were developed. The meta-evaluation model included sub-items which were important points of view of it as follow (eight items): propriety of planning, sufficiency of information, and propriety of evaluators in evaluation input; appropriateness of method and appropriateness of procedure in evaluation implementation; credibility of output in evaluation output; and usefulness of report and application of evaluation utilization. As the application of the meta-evaluation model, the total mean was lower than 3.0 of the median value of 2.73 (5 point Likert scale). Finally, it was suggested that the feedback of evaluation results should be more enforced to R&D program improvement particularly (evaluation utilization 2.50).

Key Words : R&D Program Evaluation, Meta-evaluation, Evaluation Indicators, Government capacity

I. Introduction

National research and development (R&D) programs have been an important means of science and technology (S&T) policy that contribute to the competitive power of a nation (Shulman, 2002). In Korea, national R&D programs were expanded in the 1990s, actually beginning from a *Teugjeong* (specification in English) R&D program by the Ministry of Science and Technology in 1982. R&D budgets have gradually grown, and the concerns in the efficiency of R&D investment have increased as well. The National R&D Performance Evaluation and Management Act was enacted in 2005 to improve the efficiency of R&D investment, by the act all national R&D programs have been assessed. The national R&D program evaluation system is divided into two types. The first is in-depth evaluation mostly by the program budget scale. The second is through meta-evaluation as self-evaluation by the ministries every three years[1].

Meta-evaluation is understood as an ends to judge usefulness of evaluation to monitor evaluation processes as well as to extend evaluation of evaluation systems (Dror, 1971: 3). Meta-evaluation can be classified in terms of the functionality as both formative meta-evaluation to guide evaluation, and summative meta-evaluation to report strengths and weaknesses of evaluation (stufflebeam, 1981: 151), meta-evaluation of national R&D programs – conducted by MSIP(Ministry of S&T, ICT and Future Planning) – means the latter.

There are generally two approaches in the analysis of meta-evaluation which highlight 1) critical components of evaluation values (Stufflebeam, 1974; Joint Committee, 1994), and 2) aspects of evaluation processes (Hong, 2000; Yi, 2003; Park, 2003; Hong, 2007; Hwang, 2008; Kim, 2009; Ryu, 2009). The first approach to emphasize evaluation values may not be useful to draw actual practical issues and adapt any progresses or alternatives to be suggested into meta-evaluation practices (Ryu, 2009:161). Due to the limitation, government has been choosing the other approach focusing on aspects of evaluation processes.

In general, it is perceived very important to obtain public accountability and program

1) R&D budget was about 17.2 trillion Won in 2013. And thirty two ministries and offices conducted 570 R&D programs by the standards of examination and analysis (MSIP & KISTEP, 2014a: 4-5). Meta-evaluation for national R&D programs means evaluation on self-evaluation for these R&D programs.

improvement in evaluating input, process, output, and outcomes of policies or programs. However, it may be not easy to adapt input-output approaches in the R&D evaluation, and further more complicated influential factors should be considered in technology innovation processes (Jang, 1995:17). In the area of R&D, designing evaluation models and indices makes key roles in showing initial goals and objectives to achieve and providing concrete methods about what to do for that. Accordingly, it is required to reflect unique R&D characteristics in order to develop meta-evaluation models and indices of national R&D programs in the national level.

Until now, there have mostly been studies on meta-evaluation in the public policy aspect of meta-evaluation methodology (Stufflebeam, 1974; Larson and Berliner, 1983), evaluation guideline (Joint Committee, 1994; AEA, 1995). In Korea, meta-evaluation research of national R&D program is mainly focus on individual R&D programs (YI, 2003; Park, 2003; Hong, 2007; Hwang: 2008), and thus few studies (cf. Ryu, 2011) attempted to conduct meta-evaluation of collective R&D programs. Different from previous studies, this study has just a major concern about collectives of R&D programs, national level meta-evaluations which were not conducted from individual departments[2]. Recently, evaluation studies of national R&D program use diverse approaches including quantitative research to introduce objective evaluation methods (Cho et al., 2010), budget-evaluation connection research to obtain appropriateness of evaluation results (Shin, 2008; Lee et al., 2010), evaluation index research for performance analysis (Heo et al., 2008; Jang et al., 2014). However, very few studies exist on measurement model and criteria in the national level meta-evaluation of national R&D programs. Thus, this study can extend meta-evaluation models theoretically, and practically make contribution to establish national level meta-evaluation models.

This study examines whether the meta-evaluation currently works well and whether it develops a meta-evaluation model for national R&D programs as an evaluation tool. In order to develop a meta-evaluation model and indices for national R&D programs,

2) So far, a program or special programs have been subject to the research on meta-evaluation of R&D programs in case study. Recently, a research on meta-evaluation of R&D programs was carried out in national level (Hong, 2007). But it has different point of view from this study, because of targeting in-depths evaluation (not meta-evaluation) and focusing on improvement of in-depths evaluation. This distinction is attributed to the administrative structure of Korea which vests MSIP with authority of coordinating and distributing R&D budget particularly.

the characteristics of R&D program evaluation as a target of the meta-evaluation must first be explained by discussing the essence of R&D and policy evaluation. Second, with the review about the meta-evaluation models, this study develops a meta-evaluation model for national R&D programs, sub-items, and indices. Third, this study verifies the meta-evaluation model and indices for national R&D programs using a Delphi survey. Finally, this study applies the meta-evaluation model to 93 national R&D programs and then presents key findings and implications for national R&D programs.

II. Characteristics of R&D Program Evaluation

1. Concept of R&D Program Evaluation

R&D comprises creative works undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications (OECD, 2002). A government cannot leave all of the national R&D programs to the market because it is not feasible to distribute enough resources (especially for basic science) that are needed to achieve national R&D goals[3]. Thus, a national R&D program is an expression of national S&T policies, reflecting national strategies for S&T development (Hong and Boden, 2003).

Evaluation is the process of determining the merit, worth and value of things (Scriven, 1991: 1), and it is a key analytical procedure in all disciplined intellectual and practical endeavors (Vedung, 1997: 2). Scholars in public program evaluation, for example, suggest accountability, program improvement, and basic knowledge advancement as three evaluation purposes (Vedung, 1997; Owen and Rogers, 1999, Kim, 2003; Rho, 2006; Rossi, Lipsey and Freeman, 2004). These three purposes can be applied to National R&D program evaluation as well because it bears some similarity to public program evaluation. R&D characteristics need to be considered and incorporated in the National R&D program evaluation as follows: direction of R&D activities, allocation of

3) The justification of government intervention is based on not only undersupplying technology and the uncertainty of R&D but also imperfect capital market, external economy and national security (Kim, 1993: 6-10).

limited resources, minimization of failure risk, improvement of success possibility, and motivation of researchers (Kim, 1993; Yi, 1997). Furthermore, general problems of evaluating the R&D activity and R&D organizations are also reviewed. The problems include the uncertainties in the R&D/innovation process ant its outcomes, empirical difficulties in tracing positive and negative impacts, imputation problems due to the join input and outputs nature of the R&D innovation process, where contributions from many sources, and finally political problems in claiming credit for accomplishments (Geisler, 1994).

2. Characteristics of National R&D Program Evaluation

Recently, there has been an obvious trend of attaching importance to needs and goals, and reflecting evaluation results into R&D programs. Many countries have pursued the improvement of R&D programs and the efficiency of the activities ultimately, choosing a learning evaluation system to consider whole national R&D system rather than a general (fragmentary) evaluation system to focus project units. In Korea, the evaluation focus transferred from effectiveness analysis on project implementation to long-term social and economical influence analysis on program activities in 1990s (Yang, 2003: 12), and in the process of transformation, evaluation subjects were also expanded to R&D programs linked with S&T needs and goals more directly. A main ground to emphasize R&D efficiency at the national level is that due to an increasing R&D budget, the Korean government seeks for a technology supply strategy aggressively.

Since key logics of evaluation are intended to analyze results scientifically and to reflect the evaluation results in management process for program improvement, the feedback from evaluation results in a link with the process of coordinating programs or creating new programs required for policy decision. This process will help to improve the evaluation system more effectively[4]. An R&D program evaluation system depends on characteristics and circumstances of evaluation subjects. It is difficult to set

4) In the other way, a different evaluation approach can be chosen to analyze relevance (needs vs. goals), efficiency (inputs vs. results), effectiveness (goals vs. results), and utility and sustainability (needs vs. results), etc. (European Commission, 2004: 72-73). But the existing meta-evaluation for R&D programs in Korea is based on the viewpoints of process, evidence and results.

up the evaluation system with only a standardized model and criterion due to diversity of R&D structures and fields. Evaluation results at the project level are used as basic information in superior program levels, and the information is linked closely and shared mutually (not independently) even at different evaluation levels. In this context, R&D program levels involve different contents in various fields, and thus meta-evaluation as a means for whole national R&D system analysis is important.

There are some issues that cannot be overlooked, because of characteristics of R&D program evaluation in Korea. First, higher evaluation costs should be considered because there are many programs for national R&D meta-evaluation at the national level[5]. Second, core indices should be developed and applied. The more indices to analyze R&D programs, the harder effort (cost) are required to collect and treat materials. The efforts can not only impose heavy burdens administratively on evaluators (ministries and R&D researchers) but also bring about evaluation resistance from them, which are causes to obstruct R&D productivity ultimately. Thus, it is necessary to minimize the analyses on common factors and to use core indices matching the characteristics of R&D programs as well. And last, credibility of self-evaluation results should be underlined as important evaluation items, because the results (evaluation score and grade) are used in Financial Performance Evaluation[6]. Thus the analyses on R&D program evaluation system at the national level should involve the verification process on the credibility of self-evaluation. It is essential to establish criteria of the credibility and to develop meta-evaluation model available to re-estimate the self-evaluation results.

III. Meta-evaluation Model for National R&D Programs

1. Definition of Meta-evaluation

Meta-evaluation, called “evaluation of evaluation,” was derived from doubts about the

5) Meta-evaluation in Korea is conducted every three years, of which the subjects were 109 programs (about 5.4 trillion Won totally) in 2014 (MSIP and KISTEP, 2014b: 3).

6) According to the existing meta-evaluation for R&D programs in Korea, evaluation score and grade by meta-evaluation are supposed to reflect on Financial Performance Evaluation.

utility of policy evaluation. Historically, the concept of meta-evaluation has been expanded to evaluation of evaluation system widely from evaluation on the process narrowly. Dror (1971) expands the concept of meta-evaluation to the overall system of process, and considers meta-evaluation as evaluation of evaluation system, emphasizing the feedback system of evaluation. Meta-evaluation is a means to simply help evaluators meet their goals by providing diagnostic feedback and helpful advice about what to do (Cook and Gruder, 1978: 6). In this study, following Stufflebeam's (1981) explanation on roles of meta-evaluation, we define meta-evaluation as a series of works to examine the adequacy of the original evaluation that enhance the utilization of an evaluation and result in an improvement of the quality of the evaluation. Its scope includes not only the evaluation process but the overall evaluation system.

Among the scholars that refer to the framework of meta-evaluation, Larson and Berliner (1983) propose three components that involve technical adequacy, utility, and efficiency. Scriven (1991) suggests a key evaluation checklist for meta-evaluation that is composed of foundations, sub-evaluations, and conclusions[7]. As an evaluation guideline, the Joint Committee (1994) presents the Program Evaluation Standards, which is composed of utility, feasibility, propriety, and accuracy. The American Evaluation Association (1995) proposes systematic inquiry, competence, integrity/honesty, respect for people, and responsibilities for general and public welfare in the viewpoints of evaluation.

In case studies on R&D meta-evaluation, the components for meta-evaluation are evaluation paradigm, resources, implementation and utilization (Yi, 2003); environment, input, implementation and effectiveness (Hong, 2007); context, input, process, output (Kim, 2009); foundation, planning, input, implementation and results (Ryu, 2009). Particularly, Hong's study is related to this study because it analyzed the evaluation (in-depth) system at the national level. The existing studies present main points, namely, evaluation items, and each component, which can be arranged as 〈Table 1〉.

7) Scriven (2006) modified the components by KEC model to four steps: executive summary, preface and methodology in preliminaries (part A); background and context, description and definitions, consumers, resources, and value in foundations (part B); process, outcomes, cost, comparisons, and generalizability in subevaluation (part C); synthesis, recommendations and explanations, responsibility and justification, report and support, and metaevaluation in conclusion (part D).

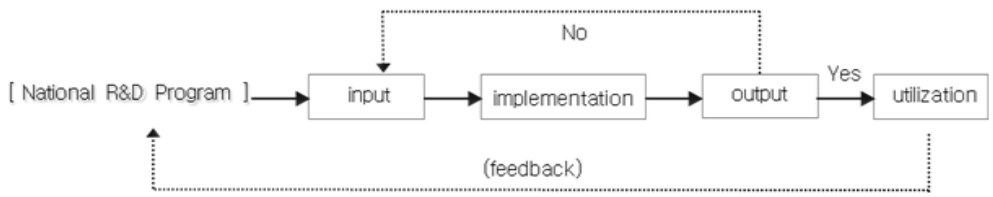
〈Table 1〉 Main points of meta-evaluation in previous research

Components	Main points	Larson & Berliner (1983)	Scriven (1991)	JC (1994)	AEA (1995)	Yi (2003)	Hong (2007)	Kim (2009)	Ryu (2009)
Input	goal	√	√	√		√	√	√	
	stakeholders	√	√	√	√				
	design	√	√	√		√			√
	institutionalization					√		√	
Implemen -tation	information	√	√	√	√	√	√	√	√
	evaluators	√	√	√	√	√	√	√	√
	criterion					√	√	√	√
	Methods	√				√	√	√	√
Output	period	√				√	√	√	√
	process		√	√	√	√	√	√	√
	communication	√		√	√			√	√
	credibility	√	√	√	√		√		
Utilization	reports		√			√	√	√	√
	distribution			√		√		√	√
	reflection	√	√	√	√	√		√	√
	utilization system			√	√		√	√	

2. Design of Meta-evaluation Model for National R&D programs

In this study, the meta-evaluation model is based on the understanding that an R&D program evaluation system is a circulation of input, implementation, output, and utilization. Generally, the program system is understood as a circulation of input, implementation, output under the regular environment, and information (i.e., feedback) about these components (Chen, 2005: 5). Program evaluation can be recognized as one of the programs as well. Thus, environment, input, implementation, output, and feedback are the major components for analyzing evaluation system. Particularly, utilization as a core concept of evaluation should not be excluded (Vedung 1997; 287). Nevertheless, this model did not include the environment component because it is difficult to control many aspects in the environmental component (such as social norms, political structure, and economy). Even the items of the environmental component can be reflected through another evaluation component, like the input component[8]. Figure 1 presents

a meta-evaluation model for national R&D program. In the model, 'Yes' and 'No' mean credibility of output. If it is not credibility (i.e., 'No'), reevaluation is required. Main points presented in existing research on the meta-evaluation and characteristics of national R&D programs mentioned at a previous chapter were taken into consideration of establishing this model.



(Figure 1) Meta-evaluation model for national R&D programs

1) Evaluation input

Evaluation input refers to human and material resources that are mobilized tangibly and intangibly and includes all actions taken before carrying out evaluation. Its influence is strong since the evaluation input has an effect on overall evaluation system including the implementation and the output of next steps. Evaluation items are propriety of planning, sufficiency of information and propriety of evaluators in the input. Propriety of planning is to analyze the evaluation goal, design and system established by ministry. Its main questions are: if the goal is proper under the meta-evaluation, if the design is feasible by reflecting stakeholders' needs and drawing up practical strategy meeting to the goal, and the system is stable (main points: goal, stakeholders, design, institutionalization). Sufficiency of information is used to estimate the utility of data for evaluation. Evaluation information is deeply implicated in the credibility of evaluation output, because it has influence on the judgment of evaluators (main point: information). And propriety of evaluators, which is an essential element in the specialty of evaluation, is used to assess their expertise and the experience needed for ensuring the quality of evaluation (main point: evaluators).

In Korea, R&D program evaluation at the national level is executed under S&T

8) For example, there are evaluation criteria, evaluators and evaluation costs in evaluation context meaningfully close to the environment by Hong (2000). But evaluation criteria can be included in the implementation (Yi, 2003; Park, 2003), evaluators in the input (Larson and Berliner, 1983; Scriven, 1991; Joint Committee, 1994; AEA, 1995; Yi, 2003) and evaluation costs in the input as well (Larson and Berliner, 1983; Scriven, 1991; Joint Committee, 1994; Yi, 2003; Park, 2003).

innovation system, and ministries carrying out R&D programs are concerned with planning and conducting self-evaluation within this integrated framework in mind. In an aspect of planning, the self-evaluation system with the goal is established under the framework of meta-evaluation, for regularity of which institutional tools are needed. Human resources are composed of evaluation managers and evaluators whom the ministry (or authorized agency) utilizes to evaluate the programs of concern. And material resources are made up of various information offered for the evaluation and evaluation budget put into evaluation. They determine the quality of self-evaluation and are important parts of the input component. But there is also the necessity to exclude evaluation budget as given-environment of the public in an aspect of meta-evaluation, considering that it is difficult for ministry to control it offered by budget ministry.

2) Evaluation implementation

Implementation of evaluation itself has influence on evaluation output directly. If a program performance without adequate actions is evaluated only according to expressed contents of the program, it will be committed a fault (Rho, 2006: 140-141). It is very important to know how to implement evaluation because evaluation activities include the evaluation process and the application of methods entailed in the process. Evaluation items are appropriateness of method and appropriateness of procedure in the implementation. Appropriateness of method is used to assess how to choose and apply a series of methods for the evaluation goal rightly (main points: standards and methods). Appropriateness of procedure is used to judge the path of evaluation and to analyze how the evaluation is progressed consistently according to the original intention with evaluators' understanding of programs (main points: period, process, and communication).

MSIP presents a guideline for self-evaluation every year, with the guidelines ministries choose and apply evaluation methods and process. In the methods, it is very important to develop and apply criteria meeting self-evaluation goals. According to some studies (Vedung, 1997; MSIP and KISTEP, 2014c), examples of program evaluation criteria include effectiveness, productivity, efficiency, classified knowledge improvement, industrial development and welfare by program characteristics, and research development, manpower, international cooperation and equipment by program purpose. In addition, evaluation patterns should take into consideration since evaluation results might be changed according to internal

or external evaluation. In the process, the evaluation should be consistently conducted by existing plans as long as there is no environmental change. Finally, various methods should be utilized for the evaluation's understanding of programs as well.

3) Evaluation Output

Evaluation output means information collected from the evaluation activity. In the evaluation, reliable and valid information are ultimate products whose credibility determines the worth of its use. If it is not credible, they must be reevaluated. Generally the purposes of program evaluation are accountability, program improvement and knowledge advancement (Vedung, 1997; Rossi, Lipsey and Freeman, 2004; Rho, 2006; Kim, 2003; Owen and Rogers, 1999). Since meta-evaluation also can achieve the purposes through feedback of evaluation results[9], credibility of the output is a core item in the evaluation output which enables to give credits to the output itself (main point: credibility).

Meta-evaluation in Korea can be defined as work that confirms and verifies the results of self-evaluation. Practically, R&D program evaluation at the national level places responsibility over system improvement, even if it is not reasonable. That is, it focuses more on coordinating grades for reflecting them into superior FPE (Financial Performance Evaluation) by assessing on contents of programs than to diagnose the self-evaluation system. Meanwhile, self-evaluation is understood as assessment of the whole of input-process-output in that it targets all general R&D programs in national level[10]. Therefore, it needs to assess on these aspects including goals, contents, system, management, and achievement etc. And, it requires inspecting output targets and indices as well as consideration of R&D program characteristics (i.e., long-term programs

9) Generally the purposes of program evaluation are accountability, program improvement and knowledge advancement (Vedung, 1997; Rossi, Lipsey and Freeman, 2004; Rho, 2006; Kim, 2003; Owen and Rogers, 1999). Meta-evaluation also can achieve the purposes through feedback of evaluation results. Recheck mentioned here is taken a new angle into consideration. That is, evaluation scores on program have got to be reflected on FPE in Korea. Therefore, if self-evaluation results come rarely to credible level, it is needed to require re-selfevaluation on programs or to coordinate self-evaluation results by virtue of meta-evaluation. The subject includes not only score but also partial information. It seems that re-selfevaluation by ministries is reasonable to clear self-evaluation system and to promote acceptance on the results from ministries. In this study, 'yes' and 'no' described at meta-evaluation model (refer to figure 1) take this logic into consideration.

10) Self-evaluation is composed of four parts, i.e. planning, management, implementation and results (MSIP and KISTEP, 2014b) which have a thread of connections with aspects of input, process and output mentioned by Chen, even if the terms are different.

supported by government).

4) Evaluation utilization

Evaluation utilization refers to the whole influence on policy activities from evaluation results (Yi, 1997: 42). The value of evaluation is realized ultimately through utilization. This has been described in terms of direct use and indirect use. In the political and administrative context, utilization is mostly supposed to occur among collective actors. In direct use the recipient is personally exposed to the evaluation - either by taking part in investigative work or by being exposed to the findings through evaluator briefings or via published literature. Indirect use occurs when the transmission is mediated through some third link in the communication process between evaluators and users (Vedung, 1997: 276-277). But we need to hold both of them in concept of utilization widely. Usefulness of the report and application of output are evaluation items in the utilization. The usefulness of the report is to judge if evaluation report meets users' needs and if it is available in a timely manner. The evaluation results that are generally included in the report should be written and reported concretely and substantially (main points: report, distribution). Application of the output is used to assess the utilization of evaluation results. The judgment of the evaluation utilization is based on the reflective approach (direct use) on programs but also the sustainable system (indirect use) for the reflection (main points: reflection, improvement, utilization system).

Under the R&D program evaluation system at the national level, self-evaluation results are expressed by reports written with guidelines[11]. They are used for meta-evaluation that assesses the propriety of programs to achieve accountability. Another purpose that it is used for is program improvement, which is achieved by political reflection such as creating new program and program interruption (or coordination) and by utilization of a system for sustainable reflection.

3. Meta-evaluation Indices for National R&D programs

In this study, the meta-evaluation model considers characteristics of R&D program

11) A self-evaluation report is accompanied with supplements such as basic materials and data. So, the concept of utilization should include the report but also the supplements.

〈Table 2〉 Meta-evaluation indices for national R&D programs (proposal)

Compo-nents	Items	Indices	Main points
input	propriety of planning	I1: Is there the goal of self-evaluation within basic frame of meta-evaluation? I2: Is there enough discussion with stakeholders when designing self-evaluation. I3: Does practical strategy meet the target of self-evaluation detail? I4: Is self-evaluation institutionalized?	goal, stake-holders, design, institutionalization
	sufficiency of information	I5: Is evaluation data available enough in quantity? I6: Does the evaluation information Include core contents that can be used for self-evaluation? I7: Is the evaluation information provided to self-evaluators in time?	information
	propriety of evaluators	I8: Are self-evaluators enough in quantity? I9: Do self-evaluators have expertise? I10: Aren't self-evaluators related to the evaluation program? I11: Do persons in operation of self-evaluation have expertise?	evaluators
implemen-tation	appropriateness of method	I1: Is there evaluation standard to achieve evaluation target? I2: Is evaluation method in quality & quantity used appropriately? I3: Is proper evaluation form (external evaluation) used?	standard, method
	appropriateness of procedure	I4: Is evaluation period given appropriately? I5: Does evaluation process maintain consistently? I6: Is there enough education for self-evaluators to understand self-evaluation? I7: Is communication among stakeholders during self-evaluation process, done smoothly?	period, process, communication
output	credibility of output	O1: Is self-evaluation for the goals & contents of program appropriate? O2: Is self-evaluation for the system of program appropriate? O3: Is self-evaluation for management and implementation of program appropriate? O4: Is evaluation for output management and utilization appropriate? O5: Is self-evaluation for the achievement vs. target appropriate? O6: Is self-evaluation for setting the output target appropriate? O7: Is self-evaluation for setting the output indices appropriate?	credibility
utilization	usefulness of report	U1: Does self-evaluation report include information that users need? U2: Is self-evaluation report explained well enough for users to understand? U3: Is self-evaluation report being distributed and reported to users in time?	report, distribution
	application of output	U4: Is evaluation output being reflected for the next year's plan of R&D program? U5: Is self evaluation output being distributed and reported to all of the decision makers who can use it? U6: Is connected system for continuous reflection of self-evaluation output appropriate? U7: Is self-evaluation output opened appropriately?	reflection, improvement, utilization system

evaluation and refers to main points of existing research and has four components, eight items, and thirty-two indices as shown in 〈Table 2〉. This model emphasizes indices by achieving performance to check the intended results and to link them with goals within the framework of whole R&D system, and underline indices for utilization to focus program improvement through the feedback of evaluation results. Because of many programs and technical diversity, this model seeks to find core indices to minimize common indices in input and implementation (excluding given-environment indices) for reducing evaluation cost. Credibility of output is developed as evaluation item specially to confirm the confidence of self-evaluation, which is an important factor for reflecting evaluation results (scores) on FPE.

4. Data Collection and Application Method

The Delphi survey was conducted as a tool for verification of meta-evaluation model and indices. The issues that panels review were R&D meta-evaluation model at the national level developed through literature review with evaluation components, items and indices. There were a total of twenty four panels. This included thirteen meta-evaluation managers and assistants belonging to the Korea Institute of S&T Evaluation and Planning (KISTEP) and eleven meta-evaluation panels in 2007. These are reflected in 〈Table 3〉.

〈Table 3〉 Panel Composition in the Delphi Survey

Class	Career	Position	Numbers
KISTEP	Managers	researchers	3
	meta-evaluation assistants	researchers	10
Participants of meta-evaluation	meta-evaluation panels in 2007	researchers	2
	meta-evaluation panels in 2007	professors	9

The panels determined the importance of the components with a cutoff value of evaluation items and indices, 3.0[12] and the weight of evaluation items separated basis

12) We informed panels in advance that the items and indices less than 3.0 are supposed to be removed.

and application/development field[13]. By interviews and e-mails, the opinions of panels were synthetically arranged by median, percentile, and mean, which were provided back to the panels[14]. The survey was carried out in three rounds[15].

In the application of the meta-evaluation model, the survey's subject included ninety-three programs in both the fundamental human-power field and the equipment-based field[16]. It was conducted by six experts that had already evaluated the programs through meta-evaluation in 2008. The calculation of the score is as follows: An item's score is the average value of the indices' scores (5 point Likert scale with 5 being excellent, 4 being very good, 3 being good, 2 being poor, and 1 being very poor). A component's score is calculated by the value of adding up each weighted item's score divided by total score of each item's weight on the basis of 5 points. A final score is the average value of four component's score.

〈Table 4〉 Calculation of Score in Application of Meta-evaluation Model

Score of Item (X_j)	Score of Component (Y_k)	Final Score (Z)
$X_j = \frac{\sum_{i=1}^{mr} U_i}{m}$	$Y_k = \frac{\sum_{j=1}^{nr} X_j * W_j}{\sum_{j=1}^{nr} W_j}$	$Z = \frac{\sum_{k=1}^{pr} Y_k}{P}$
U_i = score of index m = number of index	X_j = score of item W_j = weight of item N = number of item	Y_k = score of component p = number of component (p = 4)

13) R&D steps can be classified by basic research-applied research-experimental development (OECD, 2002: 30). In this study, we assumed that there are some differences between basis and application/development in the weight. The reason that application and development were merged is difficulties in drawing a line clearly between both of them having possibility in commercialization. The weight was marked on five components at first and then on the items a component by the basis of 100 points totally.

14) We chose questionnaire with open questions so that panels could comment on this model etc. and suggest the other components, items and indexes not offered.

15) The survey periods were as follow: from October 1st to 10th (1st round), from October 17th to 24th (2nd round), and from November 1st to 15th (3rd round) in 2007. All of the panels participated in the survey to 3rd round.

16) This survey periods were as follow: from June 16th to July 23th.

IV. Delphi Survey and Application on Meta-evaluation Model

1. Meta-evaluation indices in the Delphi Survey

As a result of the first Delphi survey, the importance in evaluation components and items were higher than 3.0 of the median value. Table 5 shows the results[17].

〈Table 5〉 Result of the Delphi survey

Component & Items	Round 1			Round 2			Round 3				
	Num.	Impor-tance	Weight	Num.	Impor-tance	Weight	Num.	Im-por-tance	Std. Devi-ation	Weight	Std. Devi-ation
[input]	(10)	4.00	(22.83)	(8)	3.89	(21.63)	(8)	4.01	.4100	(21.73)	2.7385
propriety of planning	3	3.88	8.25	3	4.11	7.58	3	3.99	.4539	7.52	1.4706
fullness of information	4	3.88	6.94	2	3.74	6.33	2	3.85	.5793	6.73	1.2422
propriety of Evaluator	3	4.04	7.65	3	4.20	7.70	3	4.20	.5403	7.56	1.6896
[implementation]	(7)	4.13	(22.90)	(6)	4.23	(24.75)	(6)	4.26	.4539	(24.77)	3.2302
appropriateness of method	3	4.17	12.38	3	4.24	13.25	3	4.41	.5659	13.25	2.7386
Appropriateness of procedure	4	3.88	10.52	3	3.96	11.50	3	4.02	.4498	11.50	1.3471
[output] credibility of output	(5)	4.38	(32.40)	(5)	4.66	(30.42)	(5)	4.56	.4735	(29.23)	4.2066
credibility of planning & implementation	3	-	-	2	4.49	15.46	2	4.35	.5405	14.31	3.5990
credibility of performance	2	-	-	3	4.30	14.96	3	4.53	.4787	14.92	1.9541
[utilization]	(7)	4.04	(21.88)	(6)	4.06	(23.21)	(5)	4.09	.5166	(24.27)	2.2116
usefulness of report	3	4.08	9.46	3	4.02	10.06	2	3.87	.4806	10.75	2.9166
application of output	4	4.38	12.42	3	4.35	13.15	3	4.35	.4672	13.44	4.8353

[note] Weight values in round 1 are the average values in both basis field and application/development field.

17) Delphi panels mostly accepted four components. But some panels suggested to minimize and simplify items for considering of evaluation costs and presented main evaluation components such as equity, objectivity and consistency in worth aspect. In this study we chose the existing model, because the model of a systematic aspect can convey more obviously improvement contexts for R&D system.

In the importance of evaluation indices, two evaluation indices (less than 3.0) were removed[18]. Six evaluation indices have been merged into three[19]. One item was separated into two items and related indices were re-sorted[20]. Two evaluation indices have been added[21]. Consequently there were four components, nine items and twenty nine indices in round 1. Meanwhile in the weight of components[22], the highest output shown was 32.40 and the other components were similarly shown at around 20 points.

As a result of the 2nd Delphi survey, the importance in evaluation components and items were also higher than 3.0 of the median value. In the importance of evaluation indices, two evaluation indices were removed[23]. Four evaluation indices have been combined as two[24] and one evaluation index modified to another evaluation item[25]. Thus there were four components, nine items, and twenty-five indices in round 2. In the weight, the result of components in round 2 was similar to round 1. Last, the weight

18) 'Is self-evaluation institutionalized?' was 3.00 and 'Are self-evaluators enough in quantity?' was 2.96.

19) Merging and coordinating indices was conducted by adopting opinions of panels. There were merged in the output between 'Is self-evaluation for setting the output target appropriate?' and 'Is self-evaluation for setting the output indexes appropriate?', and between 'Is self-evaluation for the system of program appropriate?' and 'Is self-evaluation for management and implementation of program appropriate?'. And there were combined in the utilization between 'Is self evaluation output being distributed and reported to all of the decision makers who can use it?' and 'Is self-evaluation output opened appropriately?'. It is because there are lots of difficulties in separating fields of between them exactly.

20) As an existing item in the output, 'credibility of evaluation output' was divided into 'credibility of planning & implementation evaluation' and 'credibility of performance evaluation' as the newly added item. It is the result of acceptance from some panels' opinions that the item numbers are relatively small compared to the other components in the weight.

21) Added indexes were 'Was budget for self-evaluation sufficient?' in the input and 'Is connected system for continuous reflection of self- evaluation output appropriate (review & utilization of self-evaluation output)?' in the utilization, by the opinions of some panels.

22) The result in 1st round represented no significance of classification between basis and application/development so that we merged them into one from when 2nd round, according to the opinions of panels. It should be because that a program has a lot of projects which include three steps.

23) 'Is budget for self-evaluation sufficient?' was 2.94 and 'Is evaluation period given appropriately?' was 2.96.

24) There were merged in the input between 'Is evaluation data available enough in quantity?' and 'Does the evaluation information Include core contents that can be used for self-evaluation?'. And There were combined in the utilization between 'Is evaluation output being reflected for the next year's plan of R&D program?' and 'Is connected system for continuous reflection of evaluation output appropriate?' It is why they are not exclusive propositions as some panels said.

25) 'Is self-evaluation for setting the output target & indexes appropriate?' is changed from 'credibility of planning & Implementation evaluation' to 'credibility of performance evaluation' to focus on monitoring of evaluation indexes.

〈Table 6〉 Meta-evaluation indices for national R&D programs

Compo-nents	Items	Indices	Impor-tance	Std. Deviation
Input (21.7)	propriety of planning (7.5)	I1: Is there the goal of self-evaluation within basic frame of meta-evaluation?	3.60	0.6383
		I2: Is there enough discussion with stakeholders when designing self-evaluation.	4.17	0.5674
		I3: Does practical strategy meet the target of self-evaluation detail?	3.76	0.5793
	sufficiency of information (6.7)	I4: Does the evaluation information Include core contents that can be used for self-evaluation?	4.53	0.5976
		I5: Is the evaluation information provided to self-evaluators in time?	3.28	0.5403
	propriety of Evaluators (7.5)	I6: Do self-evaluators have expertise?	4.59	0.4951
		I7: Aren't self-evaluators related to the evaluation program?	4.19	0.4848
		I8: Do persons in operation of self-evaluation have expertise?	3.89	0.7516
Implemen-tation (24.8)	appropriate-ness of method (13.3)	I1: Is there evaluation standard to achieve evaluation target?	4.70	0.5335
		I2: Is evaluation method in quality & quantity used appropriately?	4.28	0.4658
		I3: Is proper evaluation form (external evaluation) used?	3.77	0.4498
	appropriate-ness of procedure (11.5)	I4: Does evaluation process maintain consistently?	4.20	0.7337
		I5: Is there enough education for self-evaluators to understand self-evaluation?	4.06	0.4699
		I6: Is communication among stakeholders during self-evaluation process, done smoothly?	3.75	0.7223
output (29.2)	credibility of planning & implementation evaluation (14.3)	O1: Is self-evaluation for the goals & contents of program appropriate?	4.49	0.6440
		O2: Is self-evaluation for the system & management of program appropriate?	3.89	0.6181
	credibility of performance evaluation (14.9)	O3: Is self-evaluation for the achievement vs. target appropriate?	4.65	0.4764
		O4: Is self-evaluation for setting the output target & indices appropriate?	3.83	0.6222
		O5: Is evaluation for output management and utilization appropriate?	4.48	0.5001
utilization (24.3)	usefulness of report (10.8)	U1: Does self-evaluation report include information that users need?	4.47	0.4806
		U2: Is self-evaluation report explained well enough for users to understand?	3.85	0.5801
	application of output (13.5)	U3: Is evaluation output being reflected for the next year's plan of R&D program?	4.72	0.4482
		U4: Is self evaluation output being distributed and reported to all of the decision makers who can use it?	3.94	0.5174
		U5: Is connected system for continuous reflection of self-evaluation output appropriate? (review and utilization of self-evaluation output)	3.84	0.6507

of items was from minimum 6.33 (fullness of information) to maximum 15.46 (credibility of planning & implementation).

Finally the result of the 3rd Delphi survey, importance in evaluation components and items were also higher than 3.0 of median value, which were verified to be appropriate. There was no index below 3.0 in the importance but two indices have been merged as one[26]. Finally, there were four components, nine items, and twenty-four indices (refer to 〈Table 6〉) in round 3. In the weight of components, the output (29.23) showed the highest and the implementation (24.77) was followed by the utilization (24.27) and the input (21.73) in the order. The weight of items ranged from a minimum of 6.73 (fullness of information) to maximum 14.92 (credibility of performance). Here the standard deviations have been reduced all over in the dimensions during three rounds.

Meanwhile, major modifications of result of Delphi survey during three rounds are shown in 〈Table 7〉.

In the application of the meta-evaluation model[27], the total mean was lower than 3.0 and the median value was 2.73 as shown in Table 8. In the scores of components, the utilization component showed the lowest at 2.50 and the other components showed similar at around 2.8 points. These results imply that self-evaluation systems need to be improved entirely and that the feedback of evaluation results should be more strictly enforced in particular with regards to R&D program improvement.

To summarize research findings, the significance of the meta-evaluation components for the national R&D projects differs by components, and such finding requires examining its relevance with the weight of the meta-evaluation components. In addition, when the median of the national R&D project's significant benchmark falls under 3, it is important to consider how to interpret such results in reality and its policy implications. For example, when the evaluation component scores less than the median,

---

26) According to the opinions of panels, there were merged in the utilization between 'Is self-evaluation report being distributed and reported to users in time?' and 'Is self evaluation output being distributed and reported to all of the decision makers who can use it?'

27) This survey turned out that Cronbach's α was over .60 in the component as the input .8426, the implementation .7893, the output .9088 and the utilization .8434 and the overall reliability was .9451 which was high enough.

〈Table 7〉 Major modifications of result of Delphi survey

Survey	Removed	Added	Combined & Modified
round 1	1) Is self-evaluation institutionalized? 2) Are self-evaluators enough in quantity?	1) Is budget for self-evaluation sufficient? 2) Is connected system for continuous reflection of self-evaluation output appropriate (review & utilization of self-evaluation output)? 3) [Evaluation item] credibility of evaluation output →[Evaluation item] credibility of evaluation output evaluation output appropriate	1) Is self-evaluation for setting the output target appropriate? + Is self-evaluation for setting the output indices appropriate? 2) Is self-evaluation for the system of program appropriate? + Is self-evaluation for management and implementation of program appropriate? 3) Is self evaluation output being distributed and reported to all of the decision makers who can use it? + Is self-evaluation output opened appropriately?
round 2	1) Is budget for self-evaluation sufficient? 2) Is evaluation period given appropriately?		1) Is evaluation data available enough in quantity? Does the evaluation information include core contents that can be used for self-evaluation? 2) Is evaluation output being reflected for the next year's plan of R&D program? Is connected system for continuous reflection of evaluation output appropriate? 3) Is self-evaluation for setting the output target & indices appropriate? (credibility of planning & Implementation evaluation → credibility of performance evaluation)
round 3			1) Is self-evaluation report being distributed and reported to users in time? Is self evaluation output being distributed and reported to all of the decision makers who can use it?

it is important to confirm whether we can interpret such finding as low significance of the evaluation component, or its relevance being low. In addition, different perceptions toward the significance of evaluation component between national R&D project's meta-evaluation and self-evaluation, along with deciphering the relevance and difference of these evaluation results, necessitate drawing theoretical implications.

2. Application of Meta-evaluation Model

In the application of the meta-evaluation model[28], the total mean was lower than 3.0 and the median value was 2.73 as shown in 〈Table 8〉. In the scores of components, the utilization component showed the lowest at 2.50 and the other components showed similar at around 2.8 points. These results imply that self-evaluation systems need to be improved entirely and that the feedback of evaluation results should be more strictly enforced in particular with regards to R&D program improvement.

All items' scores were also lower than 3.0 and especially it is required to promote the useful of report (2.32). In the concrete, each index's score was lower than 3.0

〈Table 8〉 Results of Meta-evaluation

Component & Items	Results			Final Score	Numbers of Programs			
	Weight	Mean	Std. Devi -ation		over 4	over 3	over 2	over 1
[input]	(21.7)	(2.86)	(.4906)	2.87	4	28	57	4
propriety of planning	7.5	2.92	.5858					
fullness of information	6.7	2.75	.6151					
propriety of Evaluator	7.5	2.91	.5051					
[implementation]	(24.8)	(2.83)	(.4712)	2.84	3	34	53	3
appropriateness of method	13.3	2.92	.5377					
Appropriateness of procedure	11.5	2.74	.5478					
[output] credibility of output	(29.2)	(2.71)	(.6555)	2.71	6	31	51	5
credibility of planning & implementation	14.3	2.75	.6742					
credibility of performance	14.9	2.67	.7242					
[utilization]	(24.8)	(2.48)	(.5768)	2.50	4	11	66	12
usefulness of report	10.8	2.32	.6748					
application of output	13.5	2.65	.5991					
Total	100	2.72	.4888	2.73	4	15	73	1

[note] excellent (5 point), very good (4 point), good (3 point), poor (2 point), very poor (1 point)

28) In the meta-evaluation for national level R&D programs, each index requires the score as well as opinion or recommendations for improving the self-evaluation system. But this study doesn't include the recommendations of each index in the application because of focusing to develop the meta-evaluation indices and to analyze overall self-evaluation systems in national level.

except self-evaluators' expertise (I6: 3.01), self-evaluation operators' expertise (I8: 3.04) in the input and the property of evaluation form (I3: 3.11) in the implementation. Above all, report information (U1: 2.34), report explanation (U2: 2.29) and report distribution (U4: 2.39) belonged to the lowest group.

Seventy-four programs (79.6%) among surveyed ninety-three programs was lower than 3.0. In the scores of components, there were sixty-one programs (65.6%) of input, fifty-six programs (60.2%) of implementation, fifty-six programs (60.2%) of the output, and seventy-eight programs (83.9%) of utilization all below 3.0. When we apply this meta-evaluation model, we need to re-evaluate the fifty-six programs which were below 3.0 in the output component[29]. Meanwhile, the score of fundamental human-power field (2.60) was lower than that of equipment base field (2.92).

3. Comparison to Previous Meta-evaluation Model

Comparing our Meta-evaluation Model with previous Meta-evaluation system for national R&D program, we detect difference in the method of process. Examining the process method, our model indicates that the ministry responsible for the national research development project is subject to reevaluation when the Meta-evaluation results are inappropriate. On the other hand, existing Meta-evaluation Model on national R&D program maintains that when the results are inappropriate, ministry of higher rank responsible for evaluation (MSIP) directly reevaluates the national project. Although such method of evaluation can be efficient, it could marginalize the focus on meta-evaluation by emphasizing evaluation of national research development project rather than inspection of ministry's self-evaluation system.

For these reasons, existing meta-evaluation Model is organized in a simple manner, and its main viewpoints are vastly different from our model. Existing meta-evaluation Model's components comprise three stages – process, evidence, and results – and seven indices (MSIP and KISTEP, 2014: 4). The first stage, evaluation process, includes evaluator's structure, evaluation rule, and performance analysis. In the second stage,

29) In this case, we can comprehend that it were not enough in credibility of self-evaluation on these programs, as 'No' of the output component in this meta-evaluation mode(refer to 〈figure 1〉).

evaluation evidence, evidence information, evaluation contents concreteness, and evaluation contents coincidence are included. Finally, the last stage, evaluation result, comprises self-evaluation grade. Therefore, unlike our model, we assess that the existing meta-evaluation Model and indicators for MSIP do not distinguish between evaluation components and evaluation items in terms of the weight of measurement and does not consider evaluation planning, evaluation, methods, evaluation procedures, and evaluation utilization med. One explanation for the difference is the shared perception that the existing meta-evaluation Model serves as a preliminary review process for the reevaluation of national research development project.

V. Conclusion

R&D program evaluation is used to analyze intended results scientifically and to reflect the evaluation results in the management process for program improvement. Further, its perception has been advanced from a fragmentary evaluation system focusing project units to a learning evaluation system considering the whole national R&D system. Here, meta-evaluation for national R&D programs should be established within the framework of the entire R&D system linked to S&T needs and objectives. There are expansive tools and measurement, core indices, excluding given environment indices and indices on credibility of self-evaluation which are main points in setting up the meta-evaluation model.

This meta-evaluation model considers the characteristics of R&D program evaluation and refers to main points of existing research, which has the four components (i.e. evaluation input, evaluation implementation, evaluation output, and evaluation utilization). They include important sub-items as follows: propriety of planning, sufficiency of information and propriety of evaluators in evaluation input; appropriateness of method and appropriateness of procedure in evaluation implementation; credibility of planning & implementation evaluation and credibility of output in evaluation output; and usefulness of report and application of evaluation utilization. And the items are composed of twenty-four evaluation indices. As the application of the meta-evaluation model, the

total mean of 2.73 was lower than the 3.0 median value (5 point Likert scale). Finally, we suggest that the feedback of evaluation results should be more stringently enforced, particularly for R&D program improvement (evaluation utilization 2.50).

From our research, we can draw several policy implications. First, national R&D evaluation system needs to be modified by obtaining autonomy from evaluation to serve as a foundation to boost R&D performance. Such modification is possible by transitioning the existing Meta-evaluation model based on reevaluation to function as a means for inspecting evaluation system. Recent modification of U.S. GPRA (Government Performance and Results Act) – foundation to South Korea's national R&D program evaluation system – to GPRAMA (Government Performance and Results Act Modernization Act) – can be viewed in this light, emphasizing autonomy of ministry in charge of research while inducing according responsibilities.

Second, national R&D evaluation system excessively focuses on ensuring accountability. The evaluation system needs to be modified to ensure recommendations for improving comprehensive research development policy and individual program are suggested through meta-evaluation. For this purpose, pertinent information related to R&D performance and policy needs to be collected and applied systematically during meta-evaluation process. In this regards, discontinuation of PART (performance Assessment Rating Tool) in U.S. provides several meaningful insights. Although PART succeeded in generating numerous performance related indicators, its applicability for program improvement was poor, suggesting the limitation of evaluation methods dependent on scoring.

Third, deep consideration must be given to the application of evaluation results. Because national R&D evaluation system modifies its budget based on grading, budget modification based on the findings of the evaluation is lacking. The meta-evaluation results, based on the findings of the evaluation, should serve as a feedback to determine increase in budget. In tandem, through meta-evaluation, one should be able to determine whether its findings are applied across entire program, such as in areas of program priorities, resource allocation, business methods adaptation, and modification through stakeholders.

This study is particularly meaningful for theoretically advancing the meta-evaluation

model for R&D programs that has yet to be established in Korea with an administrative structure that has vested MEST with the authority of coordinating and distributing the R&D budget. So far, some case studies on meta-evaluation of R&D programs have indicated that there is some information about programs and special programs. However, there is little research about national level programs that include various technology fields. This study has limitations to design meta-evaluation model based on special circumstance in Korea for applying and focusing practical application. Finally, we hope that future research on meta-evaluation models for national R&D programs will be continued and that science and technology will step forward by developing this kind of model.

References

American Evaluation Association (1995), *Guiding Principles for Evaluators, by Shadish, W. R. D. L. Newman, M. A. Scheirer and C. Wye*, San francisco: Jossey-Bass.

Bush, V. (1960), *Science: The Endless Frontier*, A Report to the President on a Program for Postwar Scientific Research by Vannevar Bush, Washington: National Science Foundation.

Chang, I. B. (1995), "An Analysis of Factors Influencing the Government-Sponsored R&D Performance", Thesis for the degree of PhD, Sungkyunkwan University.

Chen, H. (2005), *Practical Program Evaluation: Assessing and Improving Planning, Implementation and Effectiveness,* Thousand Oaks and London and New Delhi: Sage Publications.

Cho, W. H., Woo, J. P. and Song, C. H. (2010), "Practical Analysis of Developed Bibliometric Indicators of Evaluation of National R&D Programs: Case Study of World Class University Program", *Journal of Korea Technology Innovation Society*, 13(3): 494-512.

Cook, T. D. and Gruder, C. L. (1978), "Meta-evaluation Research", *Evaluation Quarterly*, 2(1): 5-55.

Dror, Y. (1971), *Ventures in Policy Sciences*, New York: American Elsevier Publishing.

European Commission (2004), *Evaluation EU Activities: A Practical Guide for the Commission Service,* Luxembourg: DG Budget-Evaluation Unit, July 2004.

Geisler, E. (1994), "Key Output Indicators in Performance Evaluation of Research and Development Organization", *Technological Forecasting and Social Change*, 42: 189-203.

Heo, J. G., Kim, H. D., Cho, Y. D., Cho, S. M. and Cho, S. R. (2008), "Developing Bibliometric Indicators for Analysis and Evaluation of National R&D Programs", *Journal of Korea Technology Innovation Society*, 11(3): 376-399.

Hong, H. D. (2000), "Meta-evaluation of National Large-Scale R&D Programs: A Comparison of Evaluation Systems of 6 National R&D Programs", Thesis for the degree of PhD, Policy Research in Engineering Science and Technology (PREST), University of Manchester.

Hong, H. D. and Mark, B. (2003), *R&D Program Evaluation: Theory and Practice*, England: Ashgate Publishing Ltd.

Hong, S. G. (2007), "A Study on Meta-evaluation for National R&D Evaluation System", Kookmin University.

Jang., D. H. and Kang, G. M. (2014), "A Study on the Improvement of a Patent Evaluation Indicator in Evaluating Government R&D Performance: Focusing on a Pilot Test of the Ministry of Maritime Affairs and Fisheries' Support Program", *Korean Policy Studies Review*, 23(2): 65-91.

Joint Committee on Standards for Educational Evaluation(1994), *The Program Evaluation Standards: How to Assess Evaluations of Educational Program*, Second Edition, Thousands Oaks and New Delhi and London: Sage Publications.

Kim, J. B. (1993), *Science and Technology Policy Theory*, Daeyoung Publications.

Kim, M. S. (2003), *Theory of Public Policy Evaluation*, Pakyoung Publications.

Kim, S. Y. (2009), "A Study on the Development of Meta-evaluation Indicators for Defense R&D Programs by Using FA/AHP Methods", *Korea Technology Innovation Society*, 12(1): 113-136.

Larson, R. and Berliner, L. (1983), "On Evaluating Evaluations", *Policy Sciences*, 16(2): 147-163.

Lee, J., Kim, J. S. and Kook, Y. Y. (2010), "A Study on the Process Improvement for Interoperability between National R&D Performance Evaluation and Budget Planning", *Journal of Korea Technology Innovation Society*, 13(1): 44-67.

MSIP and KISTEP (2014a), *Analysis of National R&D Programs*, Annual Report.

MSIP and KISTEP (2014b), *Evaluation Results of National R&D Programs*, Annual Report.

MSIP and KISTEP (2014c), *Guideline for Self-evaluation on National R&D Programs*.

OECD (1998), *Best Practice Guidelines for Evaluation*, Puma Policy Brief, 5(May), Public Management Service.

OECD (2002), *Frascati Manual: Standard Practice for Surveys of Research and Experimental Development,* Paris: OECD.

Owen, J. M. and Rogers, P. J. (1999), *Program Evaluation: Forms and Approaches, International edition*, London and Thousand Oaks and New Delhi: Sage Publications.

Park, J. S. (2003), "Design and Application of the Meta-evaluation for the National IT R&D Project", Thesis for the degree of PhD, Chungnam National University.

Rho, H. J. (2006), *Policy Evaluation Theory*, Bobmun Publications.

Rossi, P. H., Lipsey, M. W. and Freeman, H. E. (2004), *Evaluation: A Systematic Approach*, Seventh Edition, Thousand Oaks and London and New Delhi: Sage Publications.

Ruegg, R. and Feller, I. (2003), *A Toolkit for Evaluating Public R&D Investment: Models, Methods and Findings from ATP's First Decade,* Maryland: Economic Assessment Office, NIST.

Ryu, Y. S. (2009), "A Study on Meta-evaluation of National R&D Evaluation System in Korea", *Korean Policy Studies Review*, 18(1): 158-185.

Scriven, M. (1991), *Evaluation Thesaurus*, Fourth Edition, Newbury Park London New Delhi, Newbury Park and London and New Delhi: Sage Publications.

Scriven, M. (2006), "Key Evaluation Checklist(KEC), The Evaluation Center, Joint Committee", http://www.wmich.edu/evalctr/checklist s/kec_june06.pdf. (2014. 6.).

Shin, Y. (2008), "Relations between the Evaluation Results and Budgetary Allocations of the National R&D Programs", *Korean Journal of Policy Analysis and Evaluation*, 18(2): 151-173.

Shulman, S. (2002), *Trouble on The Endless Frontier: Science, Invention and the Erosion of the Technological Commons*, Washington, DC: New America Foundation and Public Knowledge.

Stufflebeam, D. L. (1974), *Meta Evaluation*, Occasional Paper Series, 3, Kalamazoo: The Evaluation Center, Western Michigan University.

Stufflebeam, D. L. (1981), "Meta Evaluation: Concept, Standard, and Uses", In Berk, R. A. (ed.), *Educational Evaluation Methodology: The State of the Art*, 146-163, Maryland: The Johns Hopkins University Press.

Vedung, E. (1997), *Public Policy and Program Evaluation*, New Brunswick and London: Transaction Publishers.

Yang, H. S. (2003), "A Study on Methodology of National R&D Program Evaluation and Reformation of Project Based System", Sejong University.

Yi, C. O. (1997), "Meta-evaluation for National R&D: IT Project", Thesis for the degree of PhD, Chungnam National University.

Yi, C. O. (2003), "The National R&D Evaluation System in the UK: Meta-evaluation and Applicability to the Korea Case", Thesis for the degree of PhD, Policy Research in Engineering Science and Technology(PREST), University of Manchester.

Youngsoo Ryu

earned a doctorial degree of public administration at Hanyang University in Korea and currently is a senior research fellow in Korea Institute of S&T Evaluation and Planning. His research interests are policy evaluation and performance measurement. His work has appeared in Korean Policy Sciences Review, Korean Public Administration Review, Korean Policy Studies Review, Korean Journal of Policy Analysis and Evaluation, and Korean Public Management Review.

Byungdae Choi

earned a doctorial degree of public administration at The University of Akron, U.S.A. and currently is a professor of the Department of Public Administration at Hanyang University. His research interests are urban administration, municipal administration & local autonomy, performance measurement, and government innovation. There are "A Dysfunctional Politics of American Democracy: The Case of Campaign Financing and Lobbying (The Korean Association for Governance, 2014), "A Comparative Study on Perspectives towards Vitalization of Local Council Activities (The Korean Journal of Local Government Studies, 2010)", "A Comparative Analysis on the Citizen Satisfaction Survey and Seoul Service Index in the City of Seoul (The Korean Journal of Local Government Studies, 2007), "An Empirical Study on the NGO's Roles of Policy Process: Focused on the Recognition of Related Parties (Korean Policy Studies Review, 2004)", "E-Government to Combat Corruption (International Journal of Public Administration, 2002)", "Utilization of the Citizen Evaluation System in Korean Local Government (The Korean Journal of Local Government Studies, 2000)". Also, there are "A Comparative Perspective in Local Autonomy between Korea and Japan (2010)", "Understanding of Municipal Administration (2008)", "Understanding of Governance (2002)", "Building Good Governance: Reforms in Seou l(2002)", "Urban Management in Seoul (2001)" in his books.

Sang Ok Choi

is an associate professor in the Department of Public Administration at Korea University. His current research interests focus on the study of interorganizational relationships and science and technology policy. His work has appeared in Public Administration Review, Administration Review of Public Administration, State and Local Government Review, Administration & Society, International Journal of Public Sector Management, International Journal of Emergency Management, and Journal of Homeland Security and Emergency Management.