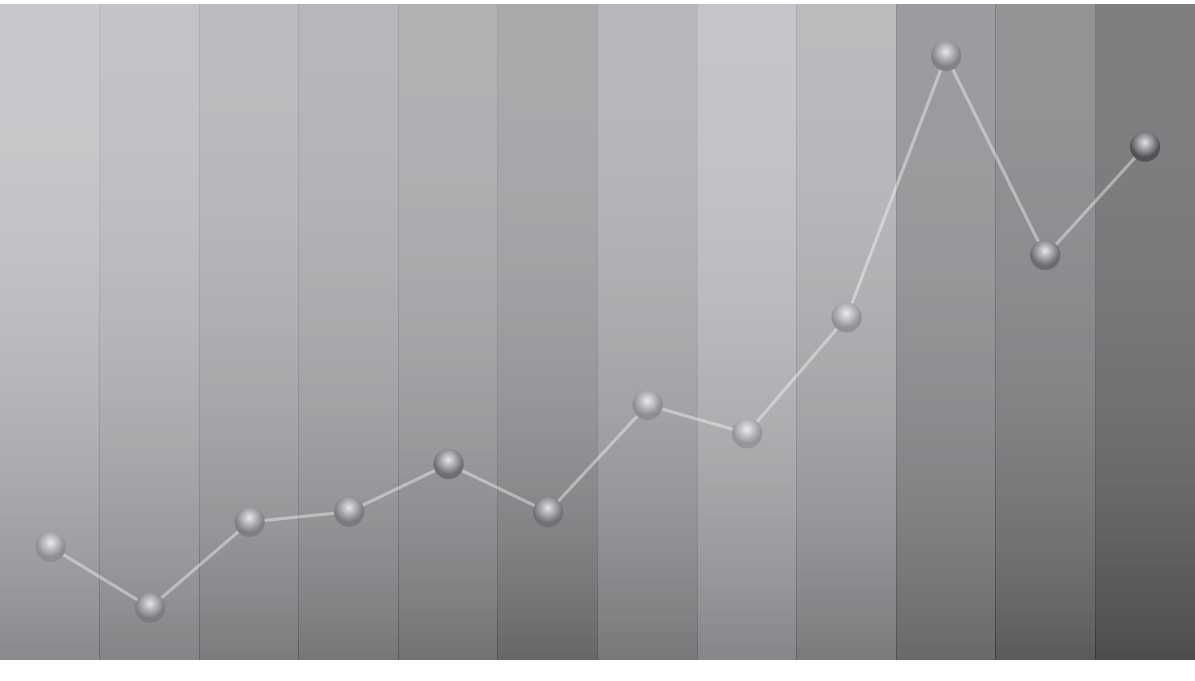
Research Paper 12–03

An Analysis of Government Programs' Performance and Its Policy Implications

December 2012

Nowook Park, Jonghak Weon



KIPF KOREA INSTITUTE OF PUBLIC FINANCE

Korea Institute of Public Finance
28, Songpa-daero 28-gil, Songpa-gu, Seoul 138-774, Korea
Tel: 82-2-2186-2114 Fax: 82-2-2186-2179
URL: www.kipf.re.kr

An Analysis of Government Programs' Performance and Its Policy Implications

December 2012

Nowook Park
Jonghak Weon

KIPF KOREA INSTITUTE OF PUBLIC FINANCE

Contents

List of Tables

List of Tables

List of Figures

I

Introduction

The purpose of this study is to analyze the performance and results of various government-funded programs, in order to infer policy implications for improving the effectiveness and efficiency of such programs. The necessity behind this study stems from the deteriorating state of national fiscal health, giving further credence to the need for improving the effectiveness and efficiency of existing budgetary programs, and overhauling the structure of tax expenditure.

South Korea's financial resources have undergone heavy exertion due to the fiscal stimulus spending aimed at reinvigorating the economy in the face of the global economic crisis, and may yet be further exhausted in the future, following the trends of a decreasing fertility rate and population aging; such circumstances necessitate an overhaul of existing budgetary programs to enhance their effectiveness and efficiency. Moreover, in cases where existing budgetary programs are high in policy value and low in effectiveness, there is a need for restructuring such programs, which are necessary for achieving policy objectives. Based on existing assessment results, it is necessary to distinguish effective program methods from ineffective ones, and to find out causes behind the ineffectiveness. Although a blanket budget cut is a common means of restoring financial stability, in many cases its effects are not sustained beyond a sufficient period of time. In order to ensure a sustainable budget retrenchment, therefore, it

is desirable to implement a differentiated budget cut by making use of performance evaluation results.

To this end, sections in which problems arise in terms of program progress and management should be identified to allow the implementation of systematic implement, incentive systems tailored for related parties, and other methods to improve program capacity. This study is aimed at finding main policy agendas by comparing and analyzing typical problems found in the respective categories of policy domain, program means, and methods of budget execution.

Many possible reasons can be attributed to the variations in the outcomes of budgetary programs depending on their characteristics: 1) certain problems may arise due to the method of implementation, 2) rapid expansion of a program may lead to complications, and 3) the performance of programs can vary depending on the line ministries / agencies. This study is intended to carry out foundational research to identify the kind of problems that arise in the sections of program planning, management, and generation of results, depending on the means of implementing programs, the environment in which programs are carried out, and the principal agent of implementation.

To date, various studies on budgetary programs have been conducted by researchers, research centers, budget authorities, and organizations including the Board of Audit and Inspection, and the National Assembly. This study, however, places emphasis on deriving issues and policy agendas related to budgetary programs by utilizing the evaluation results of previously implemented programs, rather than on analyzing the performance of specific budgetary programs. In particular, the characteristic of this study is to attempt to analyze the outcome of budgetary programs using results produced through the Self-Assessment of Budgetary Programs (SABP).

This study can be thought of as laying the foundation for follow-up studies, while also serving to test the extent to which meaningful policy agendas can be deduced from the utilization of SABP results. Therefore, this study is intended to examine the applicability of existing evaluation results for budgetary programs, while also deriving main agendas regarding budgetary programs.

The results produced by this study using performance analysis by program type can be used as the basis of policy agendas for the improvement of existing budgetary programs as well as points of concern for future programs. That is, when advancing with new budgetary programs, the issues derived from this study can be used in the deliberation process and the prioritization of programs.

In addition, this study can be used to improve the existing performance management method under the Performance Management System of Budgetary Programs (PMSBP), for which the results of this study will enable a more customized system operation reflecting the characteristics of each program. It will also allow greater focus on the core matters that ought to be examined in accordance with the different types of programs.

However, the limitation of this study lies in that it uses data from the existing PMSBP, especially the SABP, which implies that the quality of information produced from the PMSBP determines the content and usefulness of the study results. The problem is that the quality of such information inevitably depends on the capacity of the line ministries or agencies, and the capabilities of budgetary authorities to monitor and evaluate programs. In particular, the introductory phase of the PMSBP will present difficulties in producing quality information that can be utilized in a real decision-making process. It has been eight years since the PMSBP was first implemented in Korea, and there is still considerable room for improvement when it comes to the level of information related to the performance of budgetary programs. In awareness of such limitations, this study focuses on primarily analyzing information produced from the PMSBP and then deducing issues that are commonly found in the analysis by program type. Based on such issues, policy implications will be presented at the end of this paper, but full-fledged research on such policy implications will be left for follow-up studies.

The remainder of this paper is organized as follows. Section II discusses research methods and data for analysis. Section III describes South Korea's PMSBP, which provides the basis for this study's data analysis. Section IV presents the results from the data analysis and discusses policy implications. Concluding remarks and follow-up research subjects

are provided in the final section.

II

Research Scope and Method

1 Main Content of Research

The purpose of this study is to analyze the performance across budgetary programs of different types by making use of information produced from the existing PMSBP, and to deduce policy agendas.

Primarily, this study observes whether program types, according to means and purposes, have a certain impact on performance and outcome. To reiterate, it examines whether factors affecting the performance of programs vary depending on program types. Program types are classified based on methods of budget execution and relevant policy domains.

Means of budget execution can be separated into the categories of direct programs, subsidy programs, loan programs, and programs funded by contributions and investments. Direct programs are those in which central government organizations directly operate the budget and offer program services. Subsidy programs are conducted with subsidies raised or granted by the national government to provide financial aid to projects or businesses operated by bodies external to the national government. In this case, the subsidies are limited to grants given to local governments, corporations, or individuals as facility or operational funds. In this paper, private-sector subsidy programs and local-government subsidy programs will be separately analyzed. Subsidy programs can be classified into

programs funded by block grants, conditional grants, and competitive grants. Programs provided with block grants are exempted from the PMSBP by South Korean budget authorities. Loan programs direct funds to certain entities or sectors at policy interest rates below the commercial rate, by making use of public financial resources created from tax revenues, levies, and national bonds among other sources. Contribution programs refer to programs whose main source of money is government contributions, which are furnished to organizations in charge of the programs, when the legal grounds are met towards fulfilling specific purposes including the performance of national R&D projects and the accomplishment of public objectives. Investment programs refer to a form of budget supply into certain entities or sectors using public capital investments generated from tax revenues, levies, and national bonds.

Policy domains can be as diverse as follows: social welfare (for low-income earners, the handicapped, the elderly, etc.), education (primary and secondary education, higher education, vocational schools, and lifelong education), science and technology (basic and applied science), economy, administration and others. This paper, however, addresses selected areas in which there are clear issues with budgetary programs and broader availability of evaluation results. Since the detailed differentiation of policy domains poses systemic difficulties, the analysis will be carried out primarily based on the division into economy, society, and administration, while the social welfare sector will be subdivided for closer analysis.

If variations in program performance are attributable to different forms of budget execution and policy domain, the next step is to deduce policy agendas to improve the performance of programs. First, if the evaluation results exhibit meaningful differences across program types, this paper will analyze the factors behind the discrepancies. The aim of this procedure is to analyze whether there are variations in factors affecting the outcomes of programs, such as program design, performance planning, management, and results.

Throughout the analytical process, this paper will additionally observe the state of SABP evaluation, examining whether the level of difficulty in producing performance information exhibits differences depending on program type, and if evaluation results are affected by

outside factors other than the program outcome, varying across different types of programs. As such, this paper will examine if the SABP is producing meaningful performance information and deduce evaluation results with minimal bias.

Simultaneously posing a unique aspect and a level of limitation, this research analyzes program results based on the information produced by the PMSBP. Accordingly, if the information produced is poor or insufficient due to the fledgling nature of the system, it may be difficult to derive meaningful policy implications from the results of performance analysis. Although this research commissioned additional coding work to optimize the utility of data generated by the SABP, it failed to yield further practicable information.[1)] A more detailed discussion on additional efforts and the problem of limited data will follow in the main body of this paper.

This paper focuses on deducing issues and policy agendas, and selecting crucial themes from policy agendas. More specific policy alternatives will be discussed in the follow-up research. Thus, the range of this study is restricted to identifying issues by program type through the analysis of program results and deriving primary policy agendas. Full-fledged discussion on deduced policy agendas will be left for follow-up studies.

2 Research Methods and Data

The data sets used for this study were collected from SABP results produced between 2005 and 2011. The SABP evaluates one-third of the unit projects of government ministries and agencies, around 350 to 500 programs every year. The evaluation items are grouped into four categories: adequacy of program design, adequacy of performance planning, adequacy of program management, and adequacy of program results and feedback.

---

1) The Self-Assessment of Budgetary Programs checklist consists of ten plus evaluation items and each item includes subdivided criteria. Data coding was applied at the level of subdivided criteria, but substantial additional information was not found. This is thought to be related to the lack in discriminating power among evaluation items regarding program design, in which additional information was expected to be found.

- Adequacy of program design: Assessment of whether the program purpose is clear and the approach is adequate, and to discern redundancies with regards to other programs.
- Adequacy of performance planning: Assessment of whether performance indicators and performance targets are set properly.
- Adequacy of program management: Assessment of the appropriateness of budget execution and the operation of monitoring systems, and effectiveness improvement efforts.
- Program results and feedback: Assessment of the achievement of program performance targets and the feedback of evaluation results towards program improvement.

The SABP data includes nearly every budgetary program except for those whose expenditure consists mostly of personnel expenses and programs operated with a small budget. Therefore, these particular sets of data allow the most comprehensive analysis among the data currently produced by the PMSBP in South Korea.

Assessment data can be largely categorized into two groups: information on the program itself and information on assessment results generated from evaluation items. Since the program-related information included in the evaluation data was not robust, all available materials regarding the programs were used to identify their characteristics.

Although assessment results data by evaluation item are available for use, each item is under separate subdivided evaluation criteria. In particular, the sub-criteria for evaluating program design are noteworthy; for example, the sub-criteria for assessing program redundancies are designed to identify the source of overlaps, i.e., recipients, program approach, financing methods or program purpose. Thus, there is scope for subdivided evaluation criteria to provide meaningful additional information. The assessment results, however, can only be made available after undergoing an additional process of information production, reexamining the data from each ministry and agency. In this paper, additional coding was applied to the sub-criteria for evaluating program designs. Nevertheless, this process was omitted in the final paper because it generated little additional information of note, despite the extra data

coding, mainly due to shortcomings in sorting program design assessment results.

The fundamental questions regarding the performance analysis of budgetary programs are firstly whether performance varies according to program characteristics, and secondly, if so, what the cause behind the difference is. Therefore, the crucial starting point of this study is to define the characteristics of budgetary programs. In theory, the characterization can be diverse, but this study is inevitably restricted by the availability of data.

Utilizing available data, this study attempts to classify programs as follows: 1) classification by means of budget execution, 2) by policy areas, 3) by environment of implementation, and 4) by characteristics of organizations in charge of programs.

According to budget execution methods, programs are classified as direct programs, private-sector subsidy programs, local-government subsidy programs, contributions and investments, SOC (social overhead capital), and facilities, etc. Policy sections are divided into economic policy and welfare policy.

Program categorization not only specifies distinctions based on budget execution methods and relevant policy areas, but also budget sizes and the line ministries/agencies, for example, assessing the variations in evaluation results depending on the budget size. In addition, differences in performance are examined according to whether the responsible government body is a ministry or an agency.

Differences of results by program type are analyzed in the following order. Assessment results are examined for differences according to program types. In order words, program types are compared for differences in total assessment scores and ratings. This is to analyze the correlation between scores by evaluation items of SABP and program types, which serves to detect problems occurring in evaluation items according to program types. During the analysis, other factors that may affect assessment results or scores are controlled.

If assessment results differ by program type, analysis of the cause will be attempted, whereby the analysis of differences by evaluation item will primarily lead to factors that resulted in the disparity of assessment results.

This is to examine whether problems arise in the program design stage, or during implementation or management, or in the accomplishment of objectives. Causal analysis of differences in assessment results allows the detection of problems that arise in each program type, which contributes to the derivation of policy implications. However, in-depth discussion on policy implications is beyond the sphere of this study, which is limited to deducing policy implications from the empirical analysis. The proposal of fundamental and concrete policy alternatives will be addressed in follow-up studies.

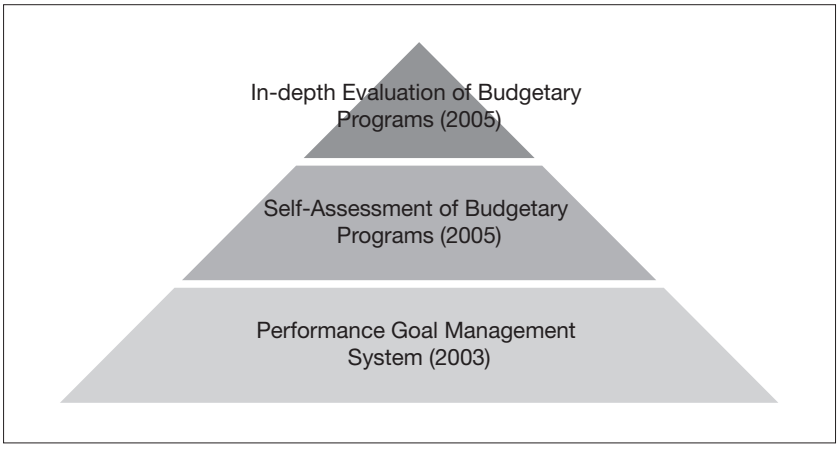
III

Overview of Performance Management System of Budgetary Programs in Korea

Prior to the discussion on data analysis, a brief description of the current state of South Korea's PMSBP will follow to aid the reader's understanding of the data. In 2003, the PMSBP was introduced in South Korea as a means of fiscal reform, along with initiatives such as the National Fiscal Management Plan (NFMP), the principle of top-down budgeting, and program budgeting. The PMSBP initiative was undertaken with an aim to supplement the monitoring function of the NFMP and to strengthen the accountability and feedback function of top-down budgeting.

Coupled with the introduction of the Performance Goal Management System, Self-Assessment of Budgetary Programs, and In-depth Evaluation of Budgetary Programs, the PMSBP has been operating as an ex-post results-based performance management and evaluation system. In terms of their individual purposes, the Performance Goal Management System was introduced for performance monitoring, Self-Assessment of Budgetary Programs for periodic program reviews, and In-depth Evaluation of Budgetary Programs for in-depth analysis of disputed programs and program groups.

[Figure III-1] Introduction of South Korea's Performance Management System of Budgetary Programs



The PMSBP plays the role of providing the basic information needed for managing the performance of budgetary programs across ministries/agencies, based on performance plans and reports submitted by government bodies. In connection with the Program Budget, the PMSBP mainly deals with performance management schemes, which consist of each ministry's strategic goals, performance objectives (programs) and management projects (unit projects), alongside corresponding performance information (performance indicators and targets).

Due to the nature of the information produced from the PMSBP, it is not easy for external individuals or organizations to utilize the PMSBP in their decision-making processes. This is because ministries/agencies often fail to produce thorough analysis despite their obligation to include target achievement analyses in the performance reports, and furthermore, even if such analysis is included, the adequacy of the analysis must be validated. In addition, it is difficult to use the PMSBP for decision making without the additional processes of analysis as the information produced through the PMSBP is not conducive to the prioritization of various budgetary programs.

In the case of the Government Performance and Result Act (GPRA)

in the U.S., after which South Korea's PMSBP was modeled, it is rare for the Office of Management and Budget (OMB) or the Congress to directly use information from performance plans and reports in the decision-making process.[2] In addition, recent global trends across budgetary performance management systems of different countries show that fiscal authorities now produce performance information that elaborates upon performance indicators and targets in order to utilize it in the decision-making process.[3] Thus it is this study's viewpoint that information produced by the PMSBP is difficult to apply to the performance analysis of budgetary programs.

The PMSBP, however, provides information that can be meaningfully utilized within ministries and agencies. In particular, the information enables senior decision-makers to carry out systematic monitoring based on schemes for performance goals and the achievement level of performance targets. In addition, through such basic monitoring, they can analyze budgetary programs' portfolios and further communicate with persons in charge of programs with regards to the causes behind the achievement or shortfall of performance targets. In essence, the PMSBP provides basic information to be used for performance management and evaluation as well as systematic information for internal management by senior decision-makers.

With regards to the PMSBP, the challenges facing South Korea concern the resolution of the discord between the program budget system and the performance goals system, and the distribution of overhead cost to budgetary programs. The revision of the National Finance Act was to mend the discrepancy between the two systems; a task that is still underway at the time of writing. In terms of the distribution of the overhead cost, which connects to gaining a better understanding of expenditure across budgetary programs, overhead cost is being distributed up to the level of programs (performance goals) in the case of special accounts and fund programs to

---

2) Based on multiple interviews conducted during 2004-2009 with officers from the OMB and the General Accounting Office (GAO) of the U.S.

3) At meetings of OECD Senior Budget Officials Network's Working Party on Performance & Results between 2011 and 2012, the key theme was that of spending reviews, which have been advocated by countries at the forefront of operating the performance management system of budgetary programs.

date. In addition, the preparation work to construe the cost of unit projects (management projects) is also underway.

The SABP was created following additional research to seek a way to utilize information produced from the PMSBP in the decision-making process. It was modeled after the U.S. Program Assessment Rating Tool (PART), and introduced as a vehicle to generate information useful for deliberation by budget authorities.

The SABP determines the priority among unit projects by systematically and periodically (once every three years) examining and rating the unit projects of each ministry/agency. Line ministries/agencies self-assess their budgetary programs, and the Ministry of Strategy and Finance (MOSF) inspects and verifies the assessment results and puts them to use in fiscal management. The legal grounds are prescribed in the sixth paragraph of Article 8 of the National Finance Act and Article 3 of its enforcement decree. The evaluation indicators are composed of 11 indicators, grouped into categories such as planning, management, results, and feedback (based on the 2011 SABP guidelines). Based on total scores combining all indicators, evaluation results are graded into a five-tier rating (very effective, effective, adequate, ineffective, and very ineffective). In principle, programs graded below ineffective are subject to consideration for a ten percent budget cut; however, their final inclusion in budget appropriation takes into account the characteristics of programs and assessment indicators as a whole.

<Table III -1> Evaluation Indicators of the 2011 SABP

Classification (scores)		Evaluation Indicators
Planning (20)	Program Design (10)	1-1. Is the program purpose clear and consistent with performance goals? 1-2. Is the program designed to eliminate redundancy or duplication of other programs? 1-3. Is the program content appropriate and advanced in an effective way?
	Performance Planning (10)	2-1. Are performance indicators clearly linked to the program purpose? 2-2. Are targets of performance indicators set in a specific and reasonable way?
Management (30)		3-1. Were efforts made for the budget to be executed as planned? 3-2. Is there a monitoring system in place to address problems? 3-3. Has effectiveness improved in achieving program goals? 3-Informatizationⓐ. Is the information system being operated and managed properly? 3-Informatizationⓑ. Were efforts made to establish a fair trading order?
Results/Feedback (50)		4-1. Have the planned targets of performance indicators been met? 4-2. Following program evaluation, is the program being carried out effectively? 4-3. Were evaluation results and comments from outsiders reflected in improving program structure?

This study analyzes whether the aforementioned evaluation items, evaluation ratings, and total scores vary depending on the characteristics of budgetary programs. Evaluation ratings and total scores reflect the overall outcomes of a program, while scores of individual evaluation items reflect the main elements that constitute the outcomes of a program. The focus of this study is to ascertain if the overall outcomes of a program and individual elements of a program demonstrate differences depending on program attributes.

The SABP process begins with the MOSF developing checklists that examine the outcomes of budgetary programs from various aspects and presenting it to ministries and agencies. Upon receiving the checklists, ministries and agencies complete questionnaires and present corresponding evidence. At the beginning of every year, the MOSF delivers to ministries and

agencies evaluation guidelines that include evaluation items and standards. The ministry also selects programs to be assessed in the corresponding year. Since 2010, unit projects with the same performance goals are assessed in the same year.

Subjects to be assessed, in principle, are selected among unit projects. Internal evaluation committees in line ministries/agencies draft primary assessment results based on materials pertaining to the SABP questionnaires, submitted by program managers. Internal evaluation committees operate and draft assessment results differently in each ministry/agency, particularly the process of compiling the opinions of committee members. An internal evaluation committee is composed of officials from the ministries/agencies and private-sector experts appointed by ministries/agencies. Following confirmation from an internal evaluation committee, the self-assessment results are conveyed to the MOSF, which is responsible for overall evaluation schemes of budgetary programs.

The MOSF re-examines evaluation results from ministries and agencies and modify parts that fail to meet assessment standards before notifying the ministries/agencies of the changes, against which they may raise objections. If ministries/agencies and the MOSF ultimately fail to reach an agreement, the end results are determined through a face-to-face meeting. The data used in this paper are the final assessment results confirmed by the MOSF.[4)]

<Table III-1> shows SABP questionnaires. Each question in the table has further subdivided criteria to be presented to ministries and agencies. For example, the question to assess whether a program is redundant or duplicative of other programs has subdivided items that examine redundancy or duplication in terms of program purposes, recipients, financing methods, program means, etc.

---

4) Some maintain that self-assessment by ministries/agencies should be respected in SABP. However, it is hard for fiscal authorities to accept on good faith alone evaluation results that only reflect opinions of ministries and agencies, since the SABP was originally introduced to prioritize programs. In fact, in the case of the PART in the U.S., departments and agencies only submit evidential materials, while the OMB carries out the actual assessment task. On this issue, officials from the OMB gave the opinion that it was expected for line departments and agencies to assign perfect scores to their own programs.

<Table III-2> Summary of SABP Results

(Unit: number of programs, %)

	Total	Very Effective	Effective	Adequate	Ineffective/ Very Ineffective
2005	555	28	100	340	87
	(100)	(5.0)	(18.0)	(61.3)	(15.7)
2006	577	30	94	388	65
	(100)	(5.2)	(16.3)	(67.2)	(11.3)
2007	585	69	143	342	31
	(100)	(11.8)	(24.4)	(58.5)	(5.3)
2008	384	11	44	226	103
	(100)	(2.9)	(11.5)	(58.9)	(26.8)
2009	346	5	14	257	70
	(100)	(1.4)	(4.0)	(74.3)	(20.2)
2010	473	0	22	335	116
	(100)	(0)	(4.7)	(70.8)	(24.5)

The In-depth Evaluation System of Budgetary Programs is to assess efficiency and adequacy of a small number of selected budgetary programs and a group of programs that are at issue, by applying program evaluation methods. At the beginning of its implementation, the evaluation was conducted by outside experts under the oversight of the competent department within the budget authorities, and the utilization of results was left at the Budget Office's discretion. As a result, the formality and binding force of the system was weak.

Recently, such operation practices have undergone some changes. The guidelines for in-depth evaluation process have been developed, the evaluation for groups of programs as well as individual programs have been formulated, and evaluation results are presented as formal agendas at the overall ministries and agencies meeting presided by the Minister of Strategy and Finance. It can be said that changes have taken place that put the focus on programs run by multiple ministries and agencies, and the evaluations have become formulized and given a stronger binding force. It is expected that such changes will raise the level of in-depth evaluation results and improve their utility.

In-depth evaluation results can be considered the most comprehensive and intensive form of assessment. However, in-depth evaluation is limited in that since only around ten programs and program groups are assessed each year, the scope of evaluation is somewhat narrow. For this reason, in-depth evaluation results are not suited for this study. However, the results of the MOSF and the National Assembly Budget Office are appropriate for use in deriving policy implications.

As such, the core of South Korea's PMSBP is comprised of the Performance Goal Management System aimed at monitoring, the SABP designed for periodic examinations, and the In-depth Evaluation of Budgetary Programs, whereby each system produces different types of information.

This study uses SABP results in analyzing performance by program type and discerning determinants of performance outcomes. This is because the information produced from the SABP is easy to analyze as it can be summarized as quantitative forms, and allows for easier comparisons between budgetary programs because the SABP analyzes programs based on common standards. Additionally, the SABP excels in its extensiveness because the system periodically assesses programs conducted by all ministries and agencies.

IV

Empirical Analysis of Self-Assessment of Budgetary Programs Results

This section examines the results of SABP over the seven years since its introduction in 2005. Considering that the main purpose of the system is to strengthen the effectiveness of programs and the efficiency of budgeting, the empirical examination of whether the SABP has been achieving its intended outcome is necessary for future improvements to the system.

Specifically, this paper will examine whether assessment results vary by program type, and if so, the assessment results of the previous seven years will provide the basis for observing the stages of origin for the differences, among planning, implementation and results.

1 Data

The data used in this paper are SABP results produced between 2005 and 2011. The number of programs to be assessed each year is shown in <Table IV-1>. The number represents the total of programs subjected to the SABP each year, excluding data that are not available in all items or scores.

In terms of the number of programs to be assessed by year, around 600 programs (555-585) were assessed during 2005-2007, but around 400

programs have been assessed since 2008. In addition, the proportion of programs to be evaluated has changed from the one-third of all programs to the one-third of programs under each purpose to examine the connectivity between unit projects and upper performance goals. By subjecting all unit projects under the same performance goals, the connectivity between unit projects and performance goals has been acquired.

<Table IV-1> Number of Programs by Year

Year	Number of Programs	Percentage
2005	555	16.77
2006	577	17.44
2007	585	17.68
2008	384	11.6
2009	346	10.46
2010	473	14.29
2011	389	11.76
Total	3,309	100

Programs subjected to the SABP are largely divided into seven categories: 1) SOC (investment) 2) facilities and equipment (large-scale facilities, purchase of equipment, and acquirement of assets) 3) other direct programs 4) contributions/investments 5) loan 6) local-government subsidy programs, and 7) private-sector subsidy programs. <Table IV-2> shows the number of the seven-type programs by year.

< Table IV-2> Number of Programs by Year and Type

Year	SOC (Investment)	Facilities & Equipment	Other Direct Programs	Contributions/ Investments	Loan	Local-government Subsidy Program	Private-sector Subsidy Program	Total
2005	54	8	144	49	68	109	123	555
2006	51	9	170	46	43	101	157	577
2007	22	6	187	69	42	103	156	585
2008	15	5	83	40	34	83	124	384
2009	7	6	112	16	23	64	118	346
2010	4	4	162	39	35	90	139	473
2011	20	0	127	25	24	53	140	389
Total	173	38	985	284	269	603	957	3,309

Among the programs, SOC, facilities and equipment, other direct programs are conducted directly by the central governmental ministries and agencies, while the remaining programs including contributions and investments, loan, local-government subsidy programs and private-sector subsidy programs are categorized as indirect programs carried out by local governments or the private sector. Among direct programs, facilities and equipment programs are not large in number, and the number of SOC programs is also on the decline. In terms of indirect programs, local-government subsidy programs and private-sector subsidy programs account for the lion's share, with contributions, investments, and loan programs making up a relatively small portion.

2 Overview of Self-Assessment of Budgetary Programs Results

The program stages to be examined in the SABP consist of three sectors that include planning (program design and performance planning), management and results. With regards to scoring, 30 points were assigned to the planning sector (15 to program design, 15 to performance planning) with 20 to the management sector, and 50 to the results sector up until 2010. Starting in 2011, 20, 30, and 50 points are respectively allotted to

planning, management, and results sectors. <Table IV-3> reports total program scores and evaluation results by sectors during 2005-2011.

<Table IV-3> State of SABP Results by Sectors

Year	Number of Programs	Total Score	Planning	Management	Results
2005	555	60.1	23.1	15.1	21.9
2006	577	59.9	22.9	14.7	22.3
2007	585	66.0	23.4	15.5	27.1
2008	384	66.6	22.4	14.4	29.5
2009	346	65.9	24.5	13.6	27.8
2010	473	62.2	23.1	16.6	23.9
2011	389	61.9	16.0	20.0	26.9

First of all, the total score edged down to 59.9 in 2006 from 60.1 in 2005, but it increased to 66 in 2007, and to 66.6 in 2008. Since 2009, however, it has been on the decline again with 65.9 in 2009, 62.2 in 2010, and 61.9 in 2011. The fact that scores showed a big shift in such sectors as planning and management in 2011 is due to the change in the allotment of scores as mentioned above.

Although not shown in the table, when converted to 100 points, the scores by stages are as follows: Program design exhibited the highest scores ranging in the 90s, management in the 70s, and performance planning in the range of the upper 50s and the lower 60s. The results sector showed the lowest scores with the range of 40-50. The reason that scores in the results sector were low in 2010 is thought to be that, with the addition of questions to examine the connectivity with upper performance goals, if upper performance goals were not met, unit projects under the same upper performance goal underwent a uniform cut in scores. Even though the score in the results sector dropped due to the change of the system in 2010, unlike other sectors, it has exhibited a steady increase in scores over the seven years from 2005 to 2011. This indicates that the achievement level of programs have improved since the introduction of the SABP, thereby notably suggesting that the system has raised the effectiveness of programs.

Ratings assigned to programs are based on the numerical scores of SABP. From 2005 to 2007, a total score of 85 or higher is rated as "effective," a range of 70 or higher to less than 85 is rated as "moderately effective," a range of 50 or higher to less than 70 rated as "adequate," and a score of less than 50 rated as "ineffective." In 2008, the four-tier ranking system expanded to a five-tier system in which a total score of 90 or higher is graded as "very effective," scores ranging from 80 to less than 90 as "effective," scores ranging from 60 to less than 80 as "adequate," the score ranging from 50 to less than 60 as "ineffective," and a score of less than 50 rated as "very ineffective."

Up until 2007, programs rated with the lowest grade of "ineffective" suffered ten percent of budget cuts, but with the expansion of the grade system in 2008, the proportion of budget cuts was divided into two steps: Programs rated as "very ineffective" are subjected to a 20 percent budget cut, programs rated as "ineffective" to a 10 percent budget cut. In addition, since 2008, programs graded as "very effective" and "effective" have been under consideration of being granted incentives in budget appropriation.[5)]

<Table IV-4> reports the state of rating based on results of SABP during 2005-2011.

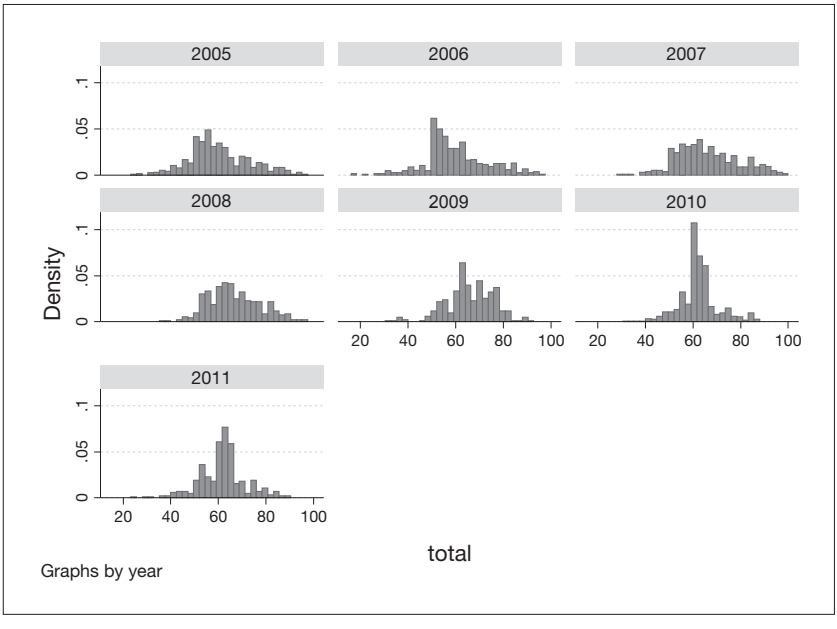
5) The percentage of incentives are not set unlike the case of budget cut, which is applied to programs rated as "ineffective" or "very ineffective."

<Table IV-4> State of Ratings (2005-2011)

Year		Very ineffective	Ineffective	Adequate	Effective (Moderately Effective)	Very effective (Effective)	Total
2005	Number of Programs	0	87	337	102	29	555
	Percentage	0	15.68	60.72	18.38	5.23	100
2006	Number of Programs	0	65	388	94	30	577
	Percentage	0	11.27	67.24	16.29	5.2	100
2007	Number of Programs	0	31	342	143	69	585
	Percentage	0	5.3	58.46	24.44	11.79	100
2008	Number of Programs	0	103	226	44	11	384
	Percentage	0	26.82	58.85	11.46	2.86	100
2009	Number of Programs	1	70	257	14	4	346
	Percentage	0.29	20.23	74.28	4.05	1.16	100
2010	Number of Programs	30	86	335	22	0	473
	Percentage	6.34	18.18	70.82	4.65	0	100
2011	Number of Programs	36	82	245	25	1	389
	Percentage	9.25	21.08	62.98	6.43	0.26	100

Features of the change in the distribution of ratings are as follows: 1) programs graded adequate appear stable in the range of 60-70 percent; 2) the proportion of programs graded moderately effective had increased every year through 2007, until a sharp decrease since 2008 when the rating system was divided into five steps; 3) even with the combination of programs rated very effective and programs rated effective since 2008, the proportion of the effective or higher ratings has dropped to about five percent since 2009; and 4) the proportion of programs rated ineffective, which had been on the decline up until 2007, become larger since 2008, when it exceeded 20 percent before surpassing 30 percent in 2011.

[Figure IV-1] Distribution of Total Score by Year



[Figure IV-1] shows the distribution of total scores of SABP results during 2005-2011. The figure indicates the distribution of scores ranging from 50 to 60 was concentrated up until 2008, while following 2008, especially after 2010, it was concentrated in the scores ranging from 60 to 70. As a whole, the distribution of scores is concentrated in the range of 50 to 70 while skewed slightly to the right in the above figure, showing the relative prevalence of programs rated as effective or very effective than those rated ineffective.

3 Evaluation Results by Program Characteristics

A. Evaluation Results by Program Type

Program types under the SABP are largely divided into direct programs and indirect programs. Direct programs are grouped into three

types such as SOC, facilities and equipment, and other direct programs, while indirect programs are grouped into four types: contributions and investments, loans, private-sector subsidy programs and local-government subsidy programs. The number of programs is presented by program type and year in <Table IV-2>, but the 2011 classification of types includes all duplicative programs, which were classified by detailed categorization.[6] While redundant programs were also present during the period of 2005-2010, those programs were included in one main type which was determined based on the allocated budget size.

The average of total scores and that of planning, management, and results, which was calculated with duplicative programs of 2011 excluded, is shown in <Table IV-5>.

<Table IV-5> SABP Results by Program Type (2005-2011)

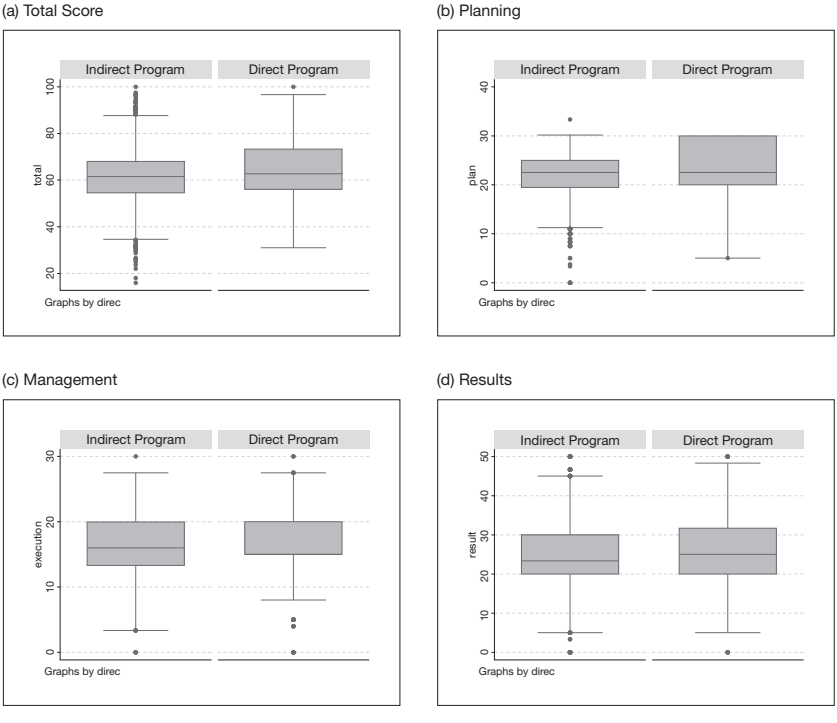
Classification	Type	Number of Samples	Total Score	Planning	Management	Results
Direct Programs	SOC (investment)	173	63.70	24.59	13.24	25.95
	Facilities & Equipment	38	65.44	25.15	15.32	25.19
	Other Direct Programs	985	65.10	23.29	16.02	26.12
Indirect Programs	Contributions / Investments	283	63.65	21.92	15.74	26.32
	Loan	269	60.95	21.37	15.57	24.26
	Local-government Subsidy Program	587	60.93	22.48	14.55	24.20
	Private-sector Subsidy Program	933	62.13	21.41	16.20	24.74

The evaluation results show that the program type with the highest total score is the facilities and equipment program, with an average over the period of 2005-2011 standing at 65.44, followed by 'other direct programs' with an average of 65.1. On the whole, the scores of direct programs were

6) As of 2011, 41 programs are included in both direct and indirect programs under the detailed categorization.

found to be higher than those of indirect programs.

[Figure IV-2] Distribution of Scores of Direct Programs and Indirect Programs



The score distributions of direct programs and indirect programs by sector show that in the planning and results sectors, similar to the total scores sector, direct programs exhibit higher scores than indirect programs. In the sector of management, however, the differences between direct and indirect programs are less pronounced. The highest average score in the management sector was earned by private-sector subsidy programs with 16.2, while SOC, despite being included in the direct program type, earned the lowest score with 13.2. That SOC exhibited the lowest score in the sector of management is largely attributed to delays in execution caused by civil complaints.

<Table IV-6> reports the results by year to highlight whether the evaluation results demonstrate yearly differences and changes.

<Table IV-6> Evaluation Results by Year and Type

Classification		2005	2006	2007	2008	2009	2010	2011
SOC	Number of Samples	54	51	22	15	7	4	20
	Total Score	62.0	65.4	65.5	65.7	69.4	64.8	58.3
	Planning	26.3	24.9	26.4	24.3	26.8	26.3	16.5
	Management	11.9	13.9	14.9	10.3	13.1	11.3	16.0
	Results	23.9	26.5	24.2	31.2	29.5	27.8	26.4
Facilities & Equipment	Number of Samples	8	9	6	5	6	4	
	Total Score	60.5	62.2	66.7	73.5	70.5	63.0	
	Planning	23.8	23.9	26.7	28.5	26.3	22.5	
	Management	14.0	15.6	16.7	14.0	13.3	20.0	
	Results	22.7	22.8	23.4	31.0	30.9	22.5	
Other Direct Programs	Number of Samples	144	170	187	83	112	162	127
	Total Score	61.7	61.1	69.0	70.5	68.4	63.4	64.4
	Planning	24.3	23.3	24.4	24.9	25.1	24.0	16.8
	Management	15.6	15.9	16.1	13.3	13.2	16.2	20.7
	Results	21.8	21.8	28.5	32.3	30.1	24.5	27.8
Contributions/ Investments	Number of Samples	49	46	69	40	16	39	25
	Total Score	59.8	57.7	69.9	65.6	63.8	63.1	62.3
	Planning	21.7	20.7	24.7	21.0	23.6	22.6	16.2
	Management	15.4	13.9	15.0	16.3	15.0	16.4	20.6
	Results	22.8	23.1	30.3	28.3	25.2	25.5	27.0
Loan	Number of Samples	68	43	42	34	23	35	24
	Total Score	58.8	58.4	63.6	65.5	64.4	59.9	58.7
	Planning	21.3	22.0	22.2	21.6	24.1	21.6	15.5
	Management	15.1	14.1	14.5	16.5	14.1	16.5	20.2
	Results	22.4	22.3	26.9	27.4	26.2	23.1	23.8
Local-government Subsidy Program	Number of Samples	109	101	103	83	64	90	53
	Total Score	58.8	56.5	62.2	66.4	64.2	61.1	60.6
	Planning	22.2	23.0	22.8	22.8	24.7	22.9	15.9
	Management	14.9	13.2	14.2	13.6	12.7	17.2	18.4
	Results	21.7	20.2	25.3	30.0	26.8	22.6	27.1
Private-sector Subsidy Program	Number of Samples	123	157	156	124	118	139	140
	Total Score	59.6	60.0	63.8	64.5	64.5	61.9	61.2
	Planning	22.4	22.6	21.8	20.6	23.7	22.8	15.2
	Management	16.2	15.0	16.2	14.9	14.2	16.7	20.4
	Results	20.9	22.3	25.7	28.0	26.6	23.8	26.6

<Table IV-5>, <Table IV-6>, and [Figure IV-2] show the differences of evaluation results between direct programs and indirect programs. The result of the t-test, which is conducted to statically validate the difference of averages of the two groups, is as follows.

First, in terms of total scores, the average of direct programs (64.91) scored higher than indirect programs (61.9) with a significance level of 0.01. The sectors of planning and results show that there are differences between the two groups at the 0.01 level of significance. However, in the sector of management, there is no statistically significant difference between the average of the direct group (15.60) and that of the indirect group (15.68).

<Table IV-7-1> T-test on Differences Between Direct Programs and Indirect Programs: Total Score

	Number of Samples	Average	Standard Error	Standard Deviation	95% Confidence Interval	
Indirect Programs	2,072	15.586	0.096	4.372	15.398	15.775
Direct Programs	1,196	15.599	0.130	4.506	15.343	15.855
Combination	3,268	15.591	0.077	4.421	15.439	15.743
Difference		-0.013	0.161		-0.327	0.302

diff = msan(0) - msan(1) t = -7.211
Ho: diff = 0 degrees of freedom = 3,266
Ha: diff < 0 Ha: diff ! = 0 Ha: dfiff > 0
Pr(T < t) = 0.000 Pr(|T| > |t|) = 0.000 Pr(T > t) = 1.000

<Table IV-7-2> T-test on Differences Between Direct Programs and Indirect Programs: Planning Section

	Number of Samples	Average	Standard Error	Standard Deviation	95% Confidence Interval	
Indirect Programs	2,072	21.779	0.119	5.418	21.545	22.012
Direct Programs	1,196	23.540	0.147	5.083	23.252	23.829
Combination	3,268	22.423	0.094	5.364	22.239	22.607
Difference		-1.762	0.192		-2.139	-1.385

diff = msan(0) - msan(1) t = -9.159
Ho: diff = 0 degrees of freedom = 3,266
Ha: diff < 0 Ha: diff ! = 0 Ha: dfiff > 0
Pr(T < t) = 0.000 Pr(|T| > |t|) = 0.000 Pr(T > t) = 1.000

<Table IV-7-3> T-test on Differences Between Direct Programs and Indirect Programs: Management

	Number of Samples	Average	Standard Error	Standard Deviation	95% Confidence Interval	
Indirect Programs	2,072	15.586	0.096	4.372	15.398	15.775
Direct Programs	1,196	15.599	0.130	4.506	15.343	15.855
Combination	3,268	15.591	0.077	4.421	15.439	15.743
Difference		-0.013	0.161		-0.327	0.302

diff = msan(0) - msan(1) t = -0.078
Ho: diff = 0 degrees of freedom = 3,266
Ha: diff < 0 Ha: diff ! = 0 Ha: dfiff > 0
Pr(T < t) = 0.469 Pr(|T| > |t|) = 0.938 Pr(T > t) = 0.531

<Table IV-7-4> T-test on Differences Between Direct Programs and Indirect Programs: Results

	Number of Samples	Average	Standard Error	Standard Deviation	95% Confidence Interval	
Indirect Programs	2,072	24.741	0.181	8.261	24.385	25.097
Direct Programs	1,196	26.068	0.262	9.074	25.553	26.583
Combination	3,268	25.227	0.150	8.590	24.932	25.521
Difference		-1.327	0.311		-1.937	-0.717

diff = msan(0) - msan(1) t = -4.264
Ho: diff = 0 degrees of freedom = 3,266
Ha: diff < 0 Ha: diff ! = 0 Ha: dfiff > 0
Pr(T < t) = 0.000 Pr(|T| > |t|) = 0.000 Pr(T > t) = 1.000

Direct programs demonstrated better evaluation results than indirect programs in all sectors except management, which yielded no meaningful difference between direct programs and indirect programs. In general, direct programs, as they are carried out directly by the government as opposed to entrusted agents, are likely to be more efficiently conducted and to outperform indirect programs. In this respect, the lack of significant difference between the two groups in the sector of management was an unexpected result.

One of the possible explanations for this result is that SOC programs are included in the category of direct programs. As mentioned before in <Table IV-5>, SOC programs are often delayed due to civil complaints

or other various factors external to the programs including budget distribution, which are tendencies considered to be the reason for lower performance in the sector of management. Aside from SOC programs, which demonstrate lower performance in management due to external factors, the comparison between direct programs and indirect programs shows that direct programs exhibited better performance than indirect programs as expected.[7)]

Meanwhile, lower evaluation results produced by indirect programs can be attributed to inefficiencies in the implementation process. Originally, the implementation of programs is commissioned to local governments or private organizations because it is impossible for the central government to directly carry out these programs or consignment is believed to ensure higher efficiency. As pointed out by Weon Jong-hak et al. (2011), however, consignment has been causing inefficiency due to the lack of expertise and segmentation of delivery systems between the public and private sectors, exacerbated by the lack of infrastructure. Such inefficiency in the delivery systems may be the reason for lower performance of indirect programs.

Evaluation results among indirect programs show that local-government subsidy programs and loan programs gained low scores, with the former in particular exhibiting low scores in such sectors as management and results, highlighting the need for remedy.

B. Evaluation Results by Program Size

As stated above, the SABP is carried out by line ministries and agencies through a self-assessment on programs, the results of which are reviewed by the MOSF. Since a ten percent budget cut is imposed on ministries / agencies rated ineffective, ministries and agencies naturally focus on the largest programs in terms of budget size, and as a result, programs with larger budgets are expected to obtain relatively higher scores. This subsection examines whether evaluation results vary by

7) The average of direct programs in the sector of management, except for SOC programs, stood at 16.00, while that of indirect programs at 15.58. Although the discrepancy is not large the two groups show statistical differences at the 0.01 significance level.

program size.

First of all, program size is divided into small-scale programs with budgets of less than five billion won, middle-scale programs with the budget ranging from five billion to less than 30 billion won, and large-scale programs with budgets of 30 billion won and over.

<Table IV-8> reports evaluation results by program size and year.

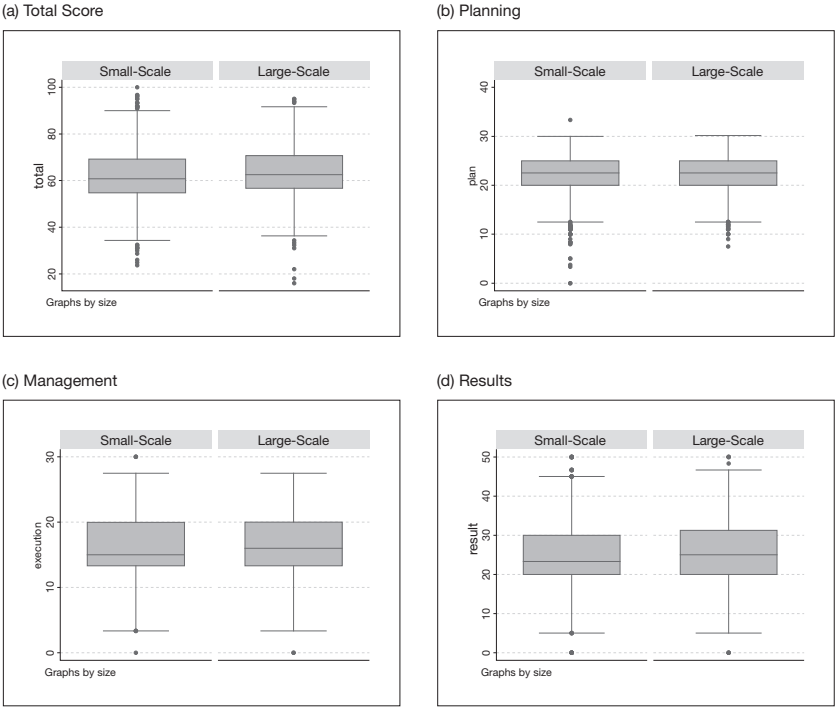
<Table IV-8> SABP Results by Program Size

		2005	2006	2007	2008	2009	2010	2011	total
Program Small-scale	Number of Samples	164	267	262	132	146	194	128	1,293
	Total Score	59.64607	58.75972	64.69499	66.59735	65.86651	60.38822	60.16711	62.06106
	Planning	22.54116	23.02491	23.27973	22.67045	24.52055	22.5201	15.45005	22.32227
	Management	15.21488	14.5621	15.85489	14.50667	13.42911	16.59072	20.23438	15.63916
	Results	21.89003	21.17245	25.56037	29.16492	27.92342	22.76193	25.85938	24.43328
Medium-scale Program	Number of Samples	228	181	192	126	122	153	112	1,114
	Total Score	58.34597	61.14635	67.2301	65.92445	65.79721	63.61164	61.70025	63.0658
	Planning	23.07105	23.19199	23.33088	21.82143	24.74385	23.49804	16.02882	22.52797
	Management	14.67	14.54724	15.20469	14.16659	13.99623	16.52682	20.0833	15.41075
	Results	20.60491	23.40702	28.69443	29.2219	27.05713	24.85476	26.48438	25.31048
Large-scale Program	Number of Samples	163	129	131	126	78	126	108	936
	Total Score	63.11902	60.5052	66.80073	67.26453	66.04013	63.39234	63.13055	64.2003
	Planning	23.53276	22.36612	23.78198	22.68452	23.97436	23.69167	16.39499	22.43969
	Management	15.69883	15.26589	15.20374	14.42397	13.38744	16.52452	19.43056	15.7516
	Results	23.87429	22.87318	27.815	30.15603	28.67833	24.63646	28.02083	26.31002

In terms of total scores by program size, small-scale programs earned 62.1, medium-scale programs 63.1 and large-scale programs 64.2. As expected, larger program sizes yielded higher evaluation results. By sector, small-scale programs in the planning sector earned 22.3, with middle-scale programs and large-scale programs gaining 22.5 and 22.4, respectively. The discrepancy between the highest and the lowest score was a mere 0.2, suggesting there is no significant difference by size in the planning sector. In

the management sector, small-scale, middle-scale, and large-scale programs earned 15.6, 15.4 and 15.8, respectively, again indicating the lack of a decisive difference between the highest score and the lowest with only 0.4 point of discrepancy. In the sector of results, however, small-scale programs gained 24.4, with medium-scale programs 25.3 and large-scale programs 26.3, showing a noticeable difference in scores. Ultimately, the difference in scores among programs divided by program size exhibited in the sector of results.

[Figure IV-3] Distribution of Scores by Program Size



<Table IV-8> and [Figure IV-3] show that the evaluation of programs distinguished by size displayed a difference between the sectors of total score and results. The following showcases the result of the t-test, which aims to validate whether the averages of small-scale programs and large-scale programs statistically different from each other.

First of all, in terms of total scores, the average of large-scale programs (64.2) was found to be higher than that of small-scale programs (62.1) at the 0.01 level of significance. In the sector of results, the averages of two groups display difference at the same level of significance as well. However, in the sectors of planning and management, no statistically significant difference was observed between the two groups, confirming that the difference between the two groups arises in the sector of results.

<Table IV-9-1> T-test on Differences Between Small-scale Programs and Large-scale Programs: Total Score

	Number of Samples	Average	Standard Error	Standard Deviation	95% Confidence Interval	
Small-scale	1,293	62.061	0.327	11.765	61.419	62.703
Large-scale	961	64.200	0.380	11.149	63.455	64.946
Combination	2,154	62.916	0.249	11.568	62.427	63.405
Difference		-2.139	0.507		-3.133	-1.145

diff = msan(0) - msan(2) t = -4.221
Ho: diff = 0 degrees of freedom = 2,152
Ha: diff < 0 Ha: diff ! = 0 Ha: dfiff > 0
Pr(T < t) = 0.000 Pr(|T| > |t|) = 0.000 Pr(T > t) = 1.000

<Table IV-9-2> T-test on Differences Between Small-scale Programs and Large-scale Programs: Planning

	Number of Samples	Average	Standard Error	Standard Deviation	95% Confidence Interval	
Small-scale	1,293	22.322	0.150	5.383	22.029	22.616
Large-scale	961	22.440	0.177	5.207	22.091	22.788
Combination	2,154	22.369	0.114	5.312	22.145	22.594
Difference		-0.117	0.234		-0.576	0.341

diff = msan(0) - msan(2) t = -0.502
Ho: diff = 0 degrees of freedom = 2,152
Ha: diff < 0 Ha: diff ! = 0 Ha: dfiff > 0
Pr(T < t) = 0.308 Pr(|T| > |t|) = 0.615 Pr(T > t) = 0.692

<Table IV-9-3> T-test on Differences Between Small-scale Programs and Large-scale Programs: Management

	Number of Samples	Average	Standard Error	Standard Deviation	95% Confidence Interval	
Small-scale	1,293	15.639	0.119	4.271	15.406	15.872
Large-scale	961	15.752	0.151	4.445	15.454	16.049
Combination	2,154	15.684	0.094	4.341	15.501	15.868
Difference		-0.112	0.191		-0.487	0.262

diff = msan(0) - msan(2) t = -0.589
Ho: diff = 0 degrees of freedom = 2,152
Ha: diff < 0 Ha: diff ! = 0 Ha: dfiff > 0
Pr(T < t) = 0.278 Pr(|T| > |t|) = 0.556 Pr(T > t) = 0.722

<Table IV-9-4> T-test on Differences Between Small-scale Programs and Large-scale Programs: Results

	Number of Samples	Average	Standard Error	Standard Deviation	95% Confidence Interval	
Small-scale	1,293	24.433	0.239	8.611	23.963	24.903
Large-scale	961	26.310	0.283	8.300	25.755	26.865
Combination	2,154	25.183	0.184	8.536	24.823	25.544
Difference		-1.877	0.373		-2.609	-1.145

diff = msan(0) - msan(2) t = -5.027
Ho: diff = 0 degrees of freedom = 2,152
Ha: diff < 0 Ha: diff ! = 0 Ha: dfiff > 0
Pr(T < t) = 0.000 Pr(|T| > |t|) = 0.000 Pr(T > t) = 1.000

The difference between small-scale programs and large-scale programs can be attributed to two factors. First, as stated above, ministries / agencies inevitably pay more attention to large-scale programs, which are subject to deep budget cuts if rated as ineffective, which is strong incentive for a more cautious management of performance. Second, there is the possibility that SABP questionnaires can be designed to be favorable to large-scale programs with particular emphasis on results; for example, inquiring whether program assessment has taken place at all.

To the second factor, regarding programs with small budgets, the additional expenditures required for program assessment add a layer of difficulty, which leads to program assessments rarely being carried out for

small-scale programs. Therefore, it can be said that the SABP is designed in a way that small-scale programs are at a relative disadvantage.

To solve this problem, the questionnaires were modified in 2010 to waive the requirement of assessment for programs with budgets of one billion won or less. However, in this research, small-size programs are defined as programs with below 2.5 billion won in budgets, and project assessment was compulsory for small-scale programs until 2009. Thus, this factor may have caused the performance gap between small-scale programs and large-scale programs.

To verify this, scores by program size to the questions of the results sector, including achievement levels of programs, implementation of project assessment, and feedback of results are described in <Table IV-10>.

<Table IV-10> Evaluation Results by Program Size: Items of Results Sector

	Achievement Level of Programs	Program Assessment	Feedback
Small-scale Program	14.55	4.84	5.28
Middle-scale Program	14.96	5.11	5.43
Large-scale Program	14.90	5.20	5.62

In terms of achievement levels, small-scale programs gained 14.55 with mid-scale programs and large-scale programs earning 14.96 and 14.90, respectively. While mid-scale programs gained the highest score, the difference between the highest score and the lowest stood at a mere 0.41. In contrast, the scores for program assessment exhibited 4.84, 5.11, and 5.20 in order of size, while the scores for feedback 5.28, 5.43, 5.62, indicating that the larger programs are, the higher the scores.

The t-test results showed that there was no significant difference between programs in terms of achievement levels. In terms of program assessment and feedback, however, there were differences at the 0.05 and the 0.01 levels of significance, respectively. Considering that the question item on feedback is designed to examine whether to fix problems found in the management process including the program assessment itself, this item can be considered to represent the level of attention line ministries / agencies

pay to corresponding programs. Therefore, the high score in the feedback sector can be interpreted to mean that line ministries / agencies are more active in their management of performance in large-scale program. Ultimately, the fact that large-scale programs outperform small-scale programs can be attributed to attention of line ministries / agencies and the question items of SABP designed more favorably for large-scale programs.

<Table IV-11-1> T-test on Differences in Achievement Levels by Program Size

	Number of Samples	Average	Standard Error	Standard Deviation	95% Confidence Interval	
Small-scale	1,277	14.553	0.215	7.677	14.132	14.975
Large-scale	848	14.903	0.262	7.636	14.389	15.418
Combination	2,125	14.693	0.166	7.661	14.367	15.019
Difference		-0.350	0.339		-1.015	0.316

diff = msan(0) - msan(2) t = -1.031
Ho: diff = 0 degrees of freedom = 2,123
Ha: diff < 0 Ha: diff ! = 0 Ha: dfiff > 0
Pr(T < t) = 0.151 Pr(|T| > |t|) = 0.303 Pr(T > t) = 0.849

<Table IV-11-2> T-test on Differences in Program Assessment by Program Size

	Number of Samples	Average	Standard Error	Standard Deviation	95% Confidence Interval	
Small-scale	743	4.837	0.133	3.631	4.576	5.099
Large-scale	467	5.198	0.163	3.533	4.876	5.519
Combination	1,210	4.976	0.103	3.597	4.773	5.179
Difference		-0.360	0.212		-0.777	0.056

diff = msan(0) - msan(2) t = -1.699
Ho: diff = 0 degrees of freedom = 1,208
Ha: diff < 0 Ha: diff ! = 0 Ha: dfiff > 0
Pr(T < t) = 0.045 Pr(|T| > |t|) = 0.090 Pr(T > t) = 0.955

<Table IV-11-3> T-test on Differences in Feedback by Program Size

	Number of Samples	Average	Standard Error	Standard Deviation	95% Confidence Interval	
Small-scale	1,215	5.281	0.100	3.481	5.085	5.476
Large-scale	781	5.621	0.112	3.142	5.400	5.842
Combination	1,996	5.414	0.075	3.356	5.266	5.561
Difference		-0.340	0.154		-0.642	-0.039

diff = msan(0) - msan(2) t = -2.214
Ho: diff = 0 degrees of freedom = 1,994
Ha: diff < 0 Ha: diff ! = 0 Ha: dfiff > 0
Pr(T < t) = 0.013 Pr(|T| > |t|) = 0.027 Pr(T > t) = 0.987

C. Evaluation Results by Program Area

This subsection will examine if performances of budgetary programs differ according to program area. The subjects in this study are limited to welfare programs and economic programs, which are currently receiving the most attention.

When classifying welfare programs and economic programs, the concept of welfare (quality of life) and spending was considered according to COFOG (Classification of the Functions of Government) defined by the United Nations, which divides government functions into ten categories.[8] However, due to the difficulty of applying this standard evenly to nearly 3,300 programs (targets of evaluation for the period of seven years), welfare programs are defined as those implemented by welfare-related ministries and agencies, and economic programs by economy-related ministries and agencies. In this study, welfare programs are defined as those carried out by the Ministry of Health and Welfare, Ministry of Gender Equality and Family, Ministry of Employment and Labor, and economic programs by the Ministry of Strategy and Finance, Ministry of Education, Science and Technology, Ministry of Food, Agriculture, Forestry and Fisheries, Small & Medium Business Administration (SMBA) and Ministry of Knowledge

8) IMF GFS classifies total government expenditures by ten functions. The South Korean government combines four of the functions into "Welfare and Quality of Life", which equates to health, social protection, housing and community amenities, recreation, culture and religion.

Economy.

<Table IV-12> shows the evaluation results by program area based on the classification above.

<Table IV-12> Evaluation Results by Program Area

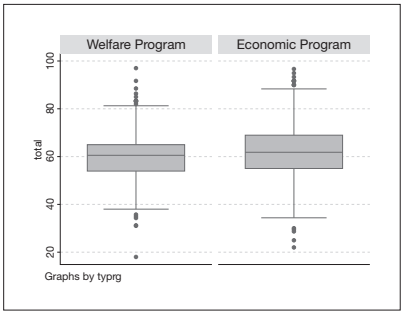
		2005	2006	2007	2008	2009	2010	2011	total
General Program	Number of Samples	410	406	358	253	224	222	153	2101
	Total Score	60.315	60.908	68.208	67.832	66.410	63.570	63.386	64.030
	Planning	23.391	23.616	24.322	22.849	24.961	23.964	16.408	23.242
	Management	14.930	14.747	15.800	13.493	13.229	16.916	19.905	15.273
	Results	21.995	22.545	28.087	31.002	28.220	24.095	27.925	25.673
Welfare Program	Number of Samples	66	92	138	67	58	99	67	587
	Total Score	58.157	58.374	59.628	60.781	63.972	59.876	59.368	59.839
	Planning	21.700	21.832	20.747	20.739	23.319	21.936	15.230	20.848
	Management	15.802	14.790	15.136	15.475	14.083	16.124	18.776	15.673
	Results	20.655	21.751	23.746	24.567	26.587	23.170	25.970	23.617
Economic Program	Number of Samples	79	79	89	64	64	152	128	655
	Total Score	60.832	56.488	66.988	67.793	65.761	61.809	60.580	62.484
	Planning	22.409	20.683	23.865	22.344	23.828	22.747	15.723	21.303
	Management	15.630	14.464	14.832	16.669	14.566	16.300	20.581	16.502
	Results	22.765	21.340	28.291	28.780	27.367	24.209	25.703	25.291

Over the past seven years, the number of economic programs performed (655) was higher than that of welfare programs (587), and the average of total scores of economic programs (62.5) was higher than that of welfare programs (59.8). However, both programs scored less than general programs (64). A close look at the scores of two programs by sector demonstrates that both scored particularly low in the planning sector, while welfare programs were not performed well in the result sector.

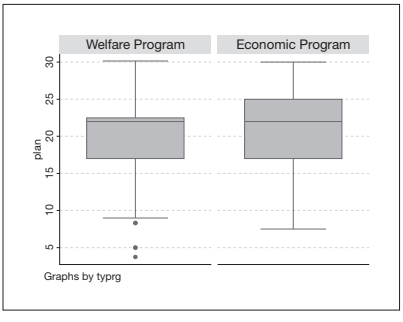
Comparing the two programs ([Figure IV-4]) shows that economic programs scored relatively better than welfare programs, especially in the planning and results sectors.

[Figure IV-4] Distribution of Scores of Welfare Programs and Economic Programs

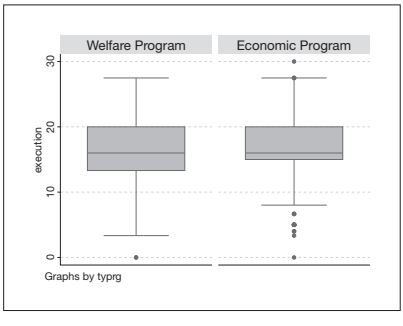
(a) Total Score



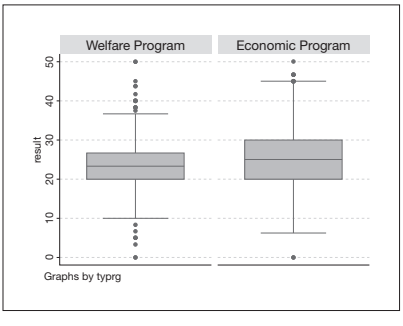
(b) Planning



(c) Management



(d) Results



<Table IV-13-1> T-test on Differences Between Welfare Programs and Economic Programs: Total Score

	Number of Samples	Average	Standard Error	Standard Deviation	95 % Confidence Interval	
Welfare Program	587	59.839	0.387	9.373	59.079	60.599
Economic Program	655	62.484	0.447	11.442	61.606	63.361
Combination	1,242	61.234	0.301	10.594	60.644	61.823
Difference		-2.645	0.598		-3.817	-1.472

diff = msan(1) - msan(2) t = -4.425
Ho: diff = 0 degrees of freedom = 1,240
Ha: diff < 0 Ha: diff ! = 0 Ha: dfiff > 0
Pr(T < t) = 0.000 Pr(|T| > |t|) = 0.000 Pr(T > t) = 1.000

<Table IV-13-2> T-test on Differences Between Welfare Programs and Economic Programs: Planning

	Number of Samples	Average	Standard Error	Standard Deviation	95 % Confidence Interval	
Welfare Program	587	20.848	0.201	4.871	20.453	21.243
Economic Program	655	21.303	0.209	5.360	20.892	21.714
Combination	1,242	21.088	0.146	5.138	20.802	21.374
Difference		-0.455	0.292		-1.027	0.118

diff = msan(1) - msan(2) t = -1.558
Ho: diff = 0 "to a significant degree"s of freedom = 1,240
Ha: diff < 0 Ha: diff ! = 0 Ha: dfiff > 0
Pr(T < t) = 060 Pr(|T| > |t|) = 0.120 Pr(T > t) = 0.940

<Table IV-13-3> T-test on Differences Between Welfare Programs and Economic Programs: Management

	Number of Samples	Average	Standard Error	Standard Deviation	95 % Confidence Interval	
Welfare Program	587	15.673	0.181	4.387	15.318	16.029
Economic Program	655	16.501	0.171	4.375	16.166	16.837
Combination	1,242	16.110	0.125	4.399	15.865	16.355
Difference		-0.828	0.249		-1.317	-0.340

diff = msan(1) - msan(2) t = -3.327
Ho: diff = 0 degrees of freedom = 1,240
Ha: diff < 0 Ha: diff ! = 0 Ha: dfiff > 0
Pr(T < t) = 0.000 Pr(|T| > |t|) = 0.000 Pr(T > t) = 1.000

<Table IV-13-4> T-test on Differences Between Welfare Programs and Economic Programs: Results

	Number of Samples	Average	Standard Error	Standard Deviation	95 % Confidence Interval	
Welfare Program	587	23.617	0.276	6.684	23.075	24.159
Economic Program	655	25.291	0.309	7.917	24.683	25.898
Combination	1,242	24.500	0.210	7.404	24.087	24.912
Difference		-1.674	0.418		-2.494	-0.853

diff = msan(1) - msan(2) t = -4.002
Ho: diff = 0 degrees of freedom = 1,240
Ha: diff < 0 Ha: diff ! = 0 Ha: dfiff > 0
Pr(T < t) = 0.000 Pr(|T| > |t|) = 0.000 Pr(T > t) = 1.000

4 Evaluation Results by Ministries and Agencies/Committees

This subsection will examine if there are differences in scores of the SABP between ministries and agencies (committees included). <Table IV-14> and [Figure IV-5] demonstrate the number of programs carried out by ministries and agencies and the averages of total scores, planning, management and results of the programs in the SABP during the period of 2005-2011.

<Table IV-14> Annual Scores of the SABP by Miwnistries and Agencies (2005-2011)

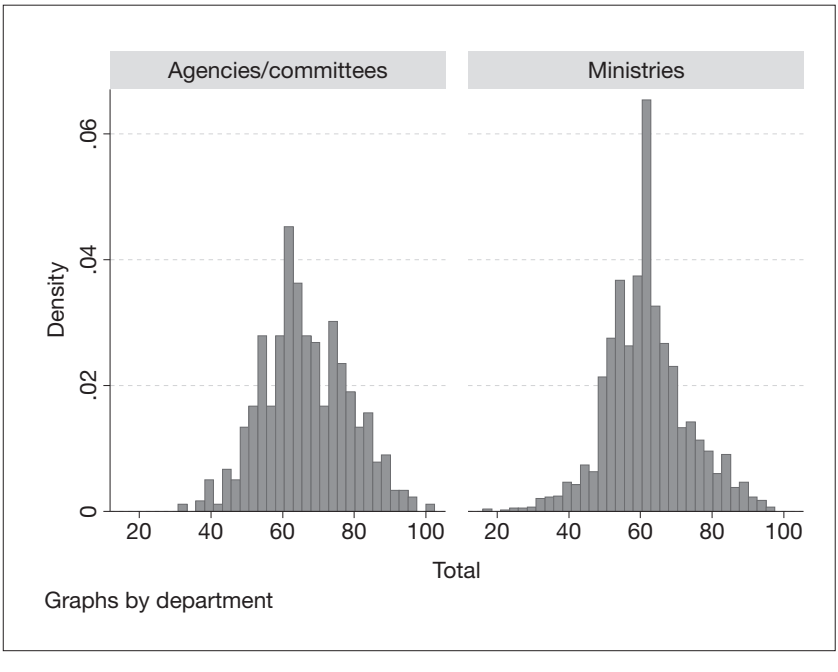
Classification		2005	2006	2007	2008	2009	2010	2011	total
Agencies/ Committees	Number of Samples	129	114	129	95	88	108	87	750
	Total Score	62.6	62.3	71.1	70.5	70.3	64.0	67.4	66.7
	Planning	23.0	23.5	24.5	22.3	25.9	23.9	16.5	22.9
	Management	15.0	15.6	16.2	15.1	13.6	17.0	21.6	16.2
	Results	24.6	23.3	30.4	31.8	30.8	24.4	29.7	27.6
Ministries	Number of Samples	426	463	456	289	258	365	302	2,634
	Total Score	59.4	59.3	64.6	65.3	64.4	61.7	60.3	61.9
	Planning	23.1	22.8	23.1	22.4	24.0	22.9	15.8	22.2
	Management	15.2	14.5	15.3	14.1	13.6	16.4	19.5	15.5
	Results	21.1	22.0	26.2	28.8	26.8	23.8	26.1	24.6

As seen in the table above, the number of programs operated by ministries (2,634) is more than three times higher than that of agencies/committees (750). However, the total score of agencies is higher than that of ministries almost every year, which shows a distinct difference in the operation methods of the two types of programs. The gap in total scores was less sizeable at two to three points during the introductory phase of the SABP across 2005 and 2006, but increased to over five points for the years following 2007 except for 2010.

The distribution of scores over seven years ([Figure IV-5]) demonstrates that ministries were generally scored within the lower 60s,

although they also received a relatively higher proportion of scores of 60 and under. While the number of programs scoring over 70 declined sharply for programs operated by ministries, more programs scoring 60 and over were operated by agencies and committees. These results are clearly seen in <Table IV-15>, which shows the distribution of grades for ministries and agencies / committees for seven years. In both ministries and agencies / committees, "adequate" accounted for the largest part with 65.9 percent and 60.6 percent, respectively. Second came ineffective for ministries with 17.2 percent, and effective for agencies with 19.8 percent. Ineffective and very ineffective accounted for 20 percent for ministries while they represented only 13 percent for agencies / committees.

[Figure IV-5] Distribution of Scores by Ministries and Agencies (Total Score)



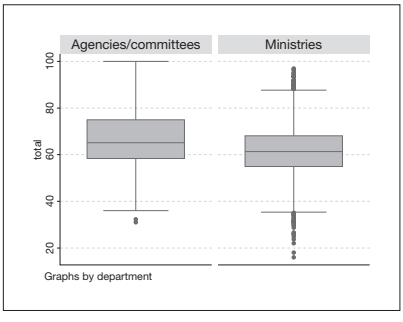
<Table IV-15> Evaluation Results by Grade by Ministries and Agencies

	Agencies/Committees	Ministries
Very Ineffective	7	54
	0.97	2.28
Ineffective	86	407
	11.86	17.17
Adequate	439	1,562
	60.55	65.91
Effective	143	264
	19.72	11.14
Very Effective	50	83
	6.9	3.5
Total	725	2,370
	100	100

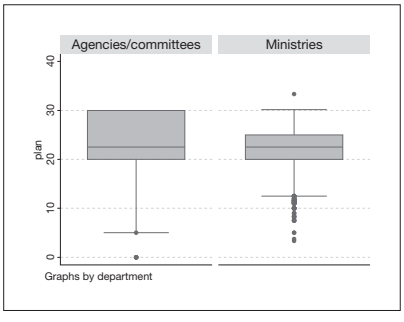
In order to find the sector of origin for the differences, the distribution of scores by sectors of planning, management, and results was examined ([Figure IV-6]). The scores in the sector of planning for agencies/committees and ministries are 16.5 and 15.8; for management 21.6 and 19.5; and for results, 29.7 and 26.1 respectively. This shows that agencies/committees scored higher in almost every sector, and surpass ministries in the number of programs scoring higher than the mode.

[Figure IV-6] Distribution of Evaluation Scores by Ministries and Agencies

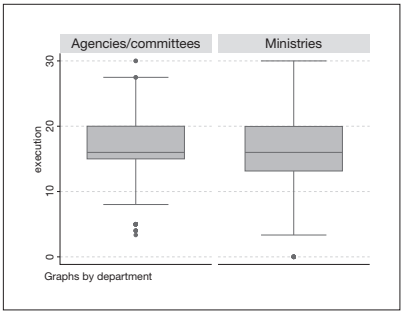
(a) Total Score



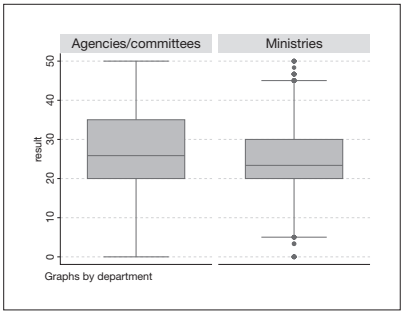
(b) Planning



(c) Management



(d) Results



A t-test was conducted to find whether the differences between the scores of the two groups were statistically significant. As expected, all the differences of averages in every sector including total scores turned out to be significant at the 0.01 significance level.

<Table IV-16-1> T-test on Differences by Ministries and Agencies: Total Score

	Number of Samples	Average	Standard Error	Standard Deviation	95 % Confidence Interval	
Agency /Committee Ministry	750 2,559	66.679 61.910	0.446 0.227	12.206 11.470	65.804 61.465	67.554 62.354
Combination	3,309	62.991	0.205	11.809	62.588	63.393
Difference		4.770	0.483		3.822	5.717

diff = msan(0) - msan(1) t = 9.868
Ho: diff = 0 degrees of freedom = 3,307
Ha: diff < 0 Ha: diff ! = 0 Ha: dfiff > 0
Pr(T < t) = 1.000 Pr(|T| > |t|) = 0.000 Pr(T > t) = 0.000

<Table IV-16-2> T-test on Differences by Ministries and Agencies: Planning

	Number of Samples	Average	Standard Error	Standard Deviation	95 % Confidence Interval	
Agency /Committee Ministry	750 2,559	22.945 22.172	0.207 0.105	5.658 5.297	22.539 21.967	23.350 22.378
Combination	3,309	22.347	0.094	5.390	22.164	22.531
Difference		0.772	0.223		0.334	1.210

diff = msan(0) - msan(1) t = 3.456
Ho: diff = 0 degrees of freedom = 3,307
Ha: diff< 0 Ha: diff ! = 0 Ha: dfiff > 0
Pr(T < t) = 0.100 Pr(|T| > |t|) = 0.001 Pr(T > t) = 0.000

<Table IV-16-3> T-test on Differences by Ministries and Agencies: Management

	Number of Samples	Average	Standard Error	Standard Deviation	95 % Confidence Interval	
Agency / Committee Ministry	750 2,559	16.203 15.492	0.160 0.088	4.394 4.461	15.888 15.319	16.518 15.665
Combination	3,309	15.653	0.077	4.455	15.501	15.805
Difference		0.711	0.185		0.349	1.073

diff = msan(0) - msan(1) t = 3.852
Ho: diff = 0 degrees of freedom = 3,307
Ha: diff < 0 Ha: diff ! = 0 Ha: dfiff > 0
Pr(T < t) = 0.100 Pr(|T| > |t|) = 0.000 Pr(T > t) = 0.000

<Table IV-16-4> T-test on Differences by Ministries and Agencies: Results

	Number of Samples	Average	Standard Error	Standard Deviation	95 % Confidence Interval	
Agency / Committee Ministry	750 2,559	27.606 24.581	0.331 0.164	9.055 8.320	26.956 24.258	28.255 24.903
Combination	3,309	25.266	0.149	8.584	24.974	25.559
Difference		3.025	0.353		2.334	3.716

diff = msan(0) - msan(1) t = 8.579
Ho: diff = 0 degrees of freedom = 3,307
Ha: diff < 0 Ha: diff ! = 0 Ha: dfiff > 0
Pr(T < t) = 1.000 Pr(|T| > |t|) = 0.000 Pr(T > t) = 0.000

<Table IV-17> accounts for the distinct differences between ministries and agencies / committees. First, in the case of agencies and committees, the proportion of direct programs to indirect programs is almost balanced for seven years while in the case of ministries, direct programs account for an overwhelming 73.05 percent. As analyzed in the previous subsection, performance management towards indirect programs was less adequate in comparison to direct programs, which verifies that program types strongly affect the results of performance management by ministries and agencies.

In addition, agencies and committees are likely to have the advantage in terms of planning programs due to the types of programs they carry out.

Because most of their programs are designed to provide services directly, they hold the advantage over ministries that are charged with more policy-oriented programs, with regards to planning, avoiding duplication, and setting up performance indicators.

<Table IV-17> Proportion of Program Types by Ministries and Agencies

	Agencies/Committees	Ministries	Total
Indirect Program	361	1,786	2,147
	(49.79)	(73.05)	(67.73)
Direct Program	364	659	1,023
	(50.21)	(26.95)	(32.27)

5 Regression Analysis on the Factors Affecting Results of the SABP

The fourth subsection observed the way in which program characteristics affected the SABP results, but in many cases, the characteristics of a particular program are interconnected. Therefore, this subsection will examine the effect of specific program characteristics with control measures over various other factors that may affect the results.

A. Total Score

This part will examine which factors affect the total SABP scores. The total score of each program was used as a dependent variable, while the type, size, and area of a program and the characteristic of ministries or agencies were used as independent variables so as to examine the impact each independent variable had on the results of budgetary programs. Besides these explanatory variables, yearly dummy variables were added to compensate for the fact that evaluation items and evaluation methods according to the items were different each year as explained in the second subsection.

As mentioned before, programs are divided into direct and indirect

programs by type; large-scale and small-scale programs by size; economic and welfare programs by area; and ministries and agencies / committees by operating organization type. The bases for the regression analysis are indirect, mid-scale, general programs operated by agencies / committees. <Table IV-18> demonstrates the estimation results.

<Table IV-18> Estimation of the SABP Results 1

	Estimated Coefficient	t	P>t
Direct Program	2.588	6.11	0.000
Small-scale Program	-1.668	-3.61	0.000
Large-scale Program	1.292	2.52	0.012
Economic Program	-0.456	-0.86	0.391
Welfare Program	-2.728	-4.88	0.000
Ministry	-3.841	-7.66	0.000
Yearly Dummy Variable	yes		
Constant	61.908	81.41	0.000
Number of Samples	3268		
Adj-Rsq	0.1059		

The estimation results show that direct programs positively affect the total scores of the SABP compared to indirect programs, and the relationship is significant when other variables such as the size or area of programs are controlled. In terms of the size of programs, small-scale programs have a negative effect and large-scale programs have a positive effect on the total scores of the SABP compared to mid-scale programs with a budget of 2.5 to 30 billion won, which suggests that the bigger the programs, the higher the evaluation results. The result also indicates that performance of large-scale programs is likely to be better managed than those of small-scale programs.

The estimated coefficients of both economic and welfare programs are negative, thus proving the result in the previous subsection, which shows evaluation results of both types of programs are worse than those of general programs. The estimation of economic programs is not statistically significant even at the 0.1 level, which means that the results of welfare

programs remained below general programs when other variables were controlled, while economic programs showed little disparity with general programs.

Evaluation results of programs carried out by ministries turned out to be very poor compared to those operated by agencies / committees, partly because indirect programs account for a large portion of ministries' programs, as mentioned in the previous subsection. In order to ascertain if the large proportion of indirect programs lead to the low evaluation results of ministries, the variable of program type was crossed with the characteristic of operating organizations. The result indicates that the interaction term of direct programs and ministries was negative at the 0.01 significance level. This result suggests that evaluation results of direct programs may suffer when performed by ministries, indicating that the large proportion of indirect programs was not necessarily the reason why the evaluation results of ministries were low.

<Table IV-19> Estimation of the SABP Results 2

	Estimated Coefficient	t	P>t
Direct Program	4.069	4.86	0.000
Small-scale Program	-1.703	-3.69	0.000
Large-scale Program	1.326	2.58	0.010
Economic Program	-0.419	-0.79	0.431
Welfare Program	-2.779	-4.97	0.000
Ministry	-2.939	-4.41	0.000
Ministry * Direct Program	-1.962	-2.05	0.041
Yearly Dummy Variable	yes		
Constant	61.908	81.41	0.000
Number of Samples	3268		
Adj-Rsq	0.1068		

B. Scores by Sector

The SABP is comprised of three major sectors: planning, management, and results. This subsection will look into which factors affect each sector using regression analysis. In this analysis, the estimators and explanatory variables used in the previous analysis on total scores were employed, but the scores for each sector instead of total scores were used as dependent variables.

1) Planning

<Table IV-20> shows the estimation results for the sector of planning.

<Table IV-20> Estimation of Factors Affecting Planning

	Estimator 1			Estimator 2		
	Estimated Coefficient	t	P>t	Estimated Coefficient	t	P>t
Direct Program	1.646	9.190	0.000	1.689	4.770	0.000
Small-scale Program	-0.362	-1.850	0.064	-0.363	-1.860	0.063
Large-scale Program	0.052	0.240	0.811	0.053	0.240	0.807
Economic Program	-0.735	-3.270	0.001	-0.734	-3.260	0.001
Welfare Program	-1.750	-7.410	0.000	-1.752	-7.410	0.000
Ministry	-0.037	-0.180	0.861	-0.011	-0.040	0.969
Ministry * Direct Programs				-0.057	-0.140	0.888
Yearly Dummy Variable	yes			yes		
Constant	22.872	80.040	0.000	22.851	71.060	0.000
Number of Samples	3268			3268		
Adj-Rsq	0.2273			0.2271		

The difference between Estimator 1 and Estimator 2 is that the interaction term of ministries and direct programs was added in Estimator 2 as an explanatory variable. The estimation results show that there is a statistically significant relationship between program planning and direct programs, economic programs or welfare programs at the 0.01 significance level. The size of programs, which was highly related to total scores, was not significantly related to planning. On the other hand, economic programs, which did not have a significant effect on total scores, turned out to have a significant negative impact on results of the planning sector, indicating that the planning sector in both economic and welfare programs fared poorly compared to general programs, which shows that the latter were more adequately planned.

The size of programs and the characteristic of operating organizations were not significantly related to planning, showing that the difference in size and operating organization type, which affected total scores, did not have an effect on planning. However, there was a significant difference between direct programs and indirect programs at the 0.01 level, suggesting that indirect programs were not as well managed as direct programs in the planning sector.

2) Management

<Table IV-21> shows the estimation results for the sector of management.

<Table IV-21> Estimation of Factors Affecting Management

	Estimator 1			Estimator 2		
	Estimated Coefficient	t		Estimated Coefficient	t	
Direct Program	-0.161	-1.04	0.296	0.947	3.11	0.002
Small-scale Program	0.154	0.92	0.359	0.129	0.77	0.444
Large-scale Program	0.278	1.49	0.137	0.303	1.62	0.105
Economic Program	0.509	2.63	0.009	0.537	2.78	0.006
Welfare Program	0.314	1.54	0.123	0.276	1.36	0.175
Ministry	-0.869	-4.76	0.000	-0.194	-0.80	0.423
Ministry * Direct Programs				-1.468	-4.21	0.000
Yearly Dummy Variable	yes			yes		
Constant	15.623	63.38	0.000	15.089	54.54	0.000
Number of Samples	3268			3268		
Adj-Rsq	0.1534			0.1577		

A notable characteristic of the estimation results in the management sector is that most of the variables were not significant, which was the case for Estimator 1, where all variables except for variables of economic programs and ministries were not significant. In the case of Estimator 2, significance was not found in all variables except for variables of direct, economic programs, and the interaction term of ministries and direct programs. In particular, direct programs, which were related to both total scores and planning at a significance level, were not significantly related to the management sector in Estimator 1 despite being significant in Estimator 2.

To sum, a significant relationship was not present between the size of programs and management. Even though differences were observed in total scores according to the size of programs, no difference was seen by size in the planning and management sectors. Based on the result, it can be assumed that differences in program size cause differences in the sector of results.

There was a statistically significant relationship between the variable of ministries and management when compared to agencies / committees in Estimator 1, but it was not the case in Estimator 2, in which the interaction term of ministries and direct programs was significantly negative, indicating that direct programs executed by ministries yielded poor results. This suggests that ministries must review their management of programs.

3) Results

<Table IV-22> shows the estimation results of factors affecting results.

<Table IV-22> Estimation of Factors Affecting Results

	Estimator 1			Estimator 2		
	Estimated Coefficient	t		Estimated Coefficient	t	
Direct Program	1.141	3.76	0.000	1.661	2.77	0.006
Small-scale Program	-1.317	-3.99	0.000	-1.330	-4.02	0.000
Large-scale Program	1.007	2.74	0.006	1.019	2.77	0.006
Economic Program	-0.109	-0.29	0.774	-0.096	-0.25	0.801
Welfare Program	-1.353	-3.38	0.001	-1.371	-3.42	0.001
Ministry	-2.650	-7.39	0.000	-2.333	-4.88	0.000
Ministry * Direct Programs				-0.690	-1.01	0.315
Yearly Dummy Variable	yes			yes		
Constant	23.825	49.22	0.000	23.574	43.29	0.000
Number of Samples	3268			3268		
Adj-Rsq	0.1352			0.1352		

The estimation for the results sector shows that the coefficient of direct programs is significant and positive, indicating that direct programs are more likely to perform better than those of indirect programs. In addition, all variables except for the variable of economic programs were significant at the 0.01 level, suggesting that explanatory variables had a more significant effect on results than on total scores.

The interaction term of ministries and direct programs was not significant in Estimator 2, indicating that ministries would not improve their performance even when they conduct direct programs. As expected, in both estimators, the size variable was significantly related to results at the 0.01 level while it was not the case with planning and management. This proves that differences in total scores depending on size arose in the result section. There can be two explanations as to why the results differ according to the size of programs. First, performance of small-scale programs may have not been well managed due to the small penalty for poor performance from those with a corresponding budget size compared to large-scale programs. Second, questions asking whether or not program assessment is being carried out in the checklist for the result sector are possibly designed unfavorably for small-scale programs.

C. Scores by Item

This part will examine which factors affect each questionnaire item, the number of which varies from 11 to 15 by year, in order to examine which factors affect the scores of each item surveyed up to date. Estimating by item may cause problems because the content and number of items are not annually consistent, and the items themselves are different each year as similar questions were combined or subdivided. Therefore, the analysis was conducted after these items were thoroughly examined with similar items merged together.

A noteworthy matter concerning scores by item is that, although the 2011 questionnaire broadly introduced the partial score system, it had previously been rarely used except when evaluating achievement levels; signifying that instead, “yes or no” items were used, and although each item was scored separately, the score by item was chosen among 0, 5, 7.5

or 10.[9] In addition, as different points were allocated to each item each year, directly applying scores as dependent variables without individual consideration is unlikely to generate meaningful results. Therefore, a probit model was used for estimation, in which "yes" is denoted as 1 and "no" as 0. However, as items associated with achievement were evaluated with partial scores from the beginning, they were estimated with evaluation scores as dependent variables, using an ordinary least squares (OLS) method.

1) Clarity of Program Purposes

The item of clarity in program purpose is designed to examine whether the purpose of a program is clear and valid.[10] In other words, the item is used to examine if the aim of a program is clear and detailed, and if the program is adequately set up to meet the performance goal stated in the performance plan. This question item was excluded from checklists in 2008 and 2009 before being reinstated in 2010; therefore, the number of samples used in the estimation is a total of 2,538, which excluded the samples from 2008 and 2009. One particular characteristic of this item is that among the answers, "no" accounts for only 1.58 percent or 40, of which 90 percent were concentrated in 2011. The reason that the answer of "no" was dominant in the 2011 evaluation is related to the evaluation of the government subsidies provided by the MOSF since 2011. The evaluation of government subsidies is designed to judge the necessity of a program funded by government subsidies, whereby poor evaluation results imply that the purpose of the program is not clear or valid. Therefore, "no" is given to the item through self-assessment.

<Table IV-23> shows the estimation results for clarity of program purposes.

---

9) Scores are different by item.

10) The item is slightly different by year. For example, the question was phrased in 2009 as "Is the purpose of the program clear and valid?" Since 2010, however, the question, "Is the purpose of the program is appropriate to achieve performance goal?" has been added to take into account the achievement of program goal.

<Table IV-23> Clarity of Program Purpose

	Estimated Coefficient	t	P>t
Direct Program	0.795	3.320	0.001
Small-scale Program	0.092	0.600	0.546
Large-scale Program	0.164	0.930	0.354
Economic Program	-0.397	-2.510	0.012
Welfare Program	-0.245	-1.410	0.158
Ministry	-0.186	-0.880	0.381
Constant	2.250	10.420	0.000
Number of Samples	2538		
Rsq	0.0815		

The estimation results for the item of clarity in program purpose show that direct programs and economic programs are in a significant relationship with the item at the 0.01 level and other variables did not have a significant effect on the item. This suggests that direct programs are likely to be set up with clearer purposes compared to indirect programs, but the result is partly attributable to the fact that programs associated with subsidies were given "no", due to the evaluation of government subsidies as mentioned before. Economic programs earned low scores in terms of clarity of program purposes compared to general programs, and there was no difference observed by size of programs or operating organization type.

2) Necessity of Government Expenditure

The item regarding the necessity of government expenditure is used to identify whether local governments and private sectors can achieve program goals without central government intervention, or if the provision of government budget or funds is inevitable. This item had been included in the checklist from 2005 to 2008, but it has been excluded since 2009. The samples used in the estimation are 2,094 and the response of "no" accounts for a mere 1.38 percent or 29.[11)]

11) The item may have been excluded from the list since 2009 because it does not bring out discriminatory results.

<Table IV-24> shows the estimation results for necessity of government spending.

<Table IV-24> Necessity of Government Spending

	Estimated Coefficient	z	P>z
Direct Program	0.072	0.440	0.657
Small-scale Program	-0.099	-0.610	0.543
Large-scale Program	0.218	0.980	0.327
Economic Program	-0.170	-0.820	0.413
Welfare Program	-0.243	-1.230	0.219
Ministry	0.228	1.300	0.195
Constant	2.085	11.020	0.000
Number of Samples	2094		
Rsq	0.0199		

No variables were significantly related to the necessity of government expenditure, suggesting the verdict that all programs in progress needed government budget or funds regardless of program type or size. However, policy direction or intention of policy makers needs to be reflected in the evaluation system to provide meaningful evaluation results. In the current evaluation system, this item does not generate meaningful results with discriminatory power.

3) Redundancy

This item examines whether programs exhibit unnecessary overlaps with other public or private programs. Any programs falling under any of the following category of cases are regarded as redundant programs: 1) if the purpose or beneficiary of a program is similar to that of other programs; 2) if there has been no adjustment process made, such as discussion between organizations conducting similar programs even when similarities between programs have been identified through external or internal evaluation; and 3) if similar programs are carried out by different entities without unavoidable reasons despite potential for integration.

This item had been included in the checklist every year from 2005 to 2011. The programs, which turned out to be similar or identical to other programs, account for 3.3 percent of the whole programs. <Table IV-25> shows the estimation results for the item of redundancy.

<Table IV-25> Redundancy

	Estimated Coefficient	z	P>z
Direct Programs	0.308	2.950	0.003
Small-scale Programs	-0.029	-0.280	0.781
Large-scale Programs	-0.024	-0.200	0.839
Economic Programs	0.245	1.720	0.085
Welfare Programs	-0.525	-4.810	0.000
Ministry	0.255	2.180	0.030
Constant	1.671	13.980	0.000
Number of Samples	3225		
Rsq	0.0512		

In this estimation, identification of redundancy denoted the dependent variable as 0, and when this was not the case, as 1. Therefore, a positive coefficient signifies the lack of redundancy. The estimation results above show that direct and welfare programs have a significant impact on the difference of redundancy at the 0.01 level. However, the coefficient of welfare programs is negative while the coefficient of ministries is positive at the 0.05 level, economic programs, positive at the 0.1 level. There was no difference by size of programs. The estimation results indicate that welfare programs are more likely to be similar to or duplicative of other programs compared to general programs.

4) Efficiency in Program Content and Process

The item of efficiency in program content and process is designed to examine whether the process, content or structure of the program is efficient and appropriate to achieve the goal of a program. The item has been included in the checklist every year except in 2010. The response

of "no" is given to the item in the following cases: 1) if the subdivided programs, which have little connectivity with the achievement of goals, are included in the program; 2) if the existing process is not as efficient as other alternatives; and 3) if the program consists of several subprograms to be evaluated, and some of the subprograms are given "no."

Over the past six years, 9.62 percent or 281 out of 2,920 programs were given "no," in terms of the item. <Table IV-26> shows the estimation results in terms of efficiency in program content and process.

<Table IV-26> Efficiency in Program Content and Process

	Estimated Coefficient	z	P>z
Direct Program	0.345	4.59	0.000
Small-scale Program	0.041	0.54	0.587
Large-scale Program	0.027	0.31	0.757
Economic Program	0.221	2.31	0.021
Welfare Program	-0.272	-3.25	0.001
Ministry	0.061	0.72	0.469
Constant	1.150	12.87	0.000
Number of Samples	2920		
Rsq	0.0254		

The estimation results about efficiency in program content and process show that there is a significant relationship between direct or welfare programs and the item at the 0.01 level, and economic programs and the item at the 0.05 significance level; however other explanatory variables were not significantly related to the item. The results show that the size of programs or characteristics of operating organizations did not have an impact on the difference of efficiency in program content and process. The estimated coefficient of welfare programs is negative, meaning that the efficiency in content and process of welfare programs was worse compared to that of other programs. Considering the results seen in the item of redundancy, the low efficiency in program content and process of

welfare programs may have been caused by redundancy.[12)]

5) Indicators in terms of Substantive Content and Connection to Program Purposes

In essence, the SABP is an evaluation system largely focused on indicators, which is why indicator-related items are central to the system even though the points allocated to indicator-related questions are not high. There are two major items about indicators: One concerning substantive content, which is designed to examine if performance goals and indicators are appropriately set up to evaluate and manage program performance, and the other is about connection, which is to check if the performance indicators are clearly connected to program goals. The former item was included in the checklist just for three years until 2007, and the latter has been included since 2005.

The programs, which were given "no" in terms of substantive content of indicators, account for only 1.69 percent or 29 out of 1,717 programs. <Table IV-27> shows the estimation results in terms of substantive content of indicators.

12) A study carried out by Weon Jong Hak et al. (2011) on efficiency in welfare programs suggests, through case studies, that the complexity and inefficiency of welfare service delivery systems result in redundancy between programs, leading to low efficiency. The result in this part can empirically support the result of the study by WeonJong Hak. Likewise, welfare programs have aspects to be examined more thoroughly than other programs. So in the following subsection, welfare programs will be analyzed in detail.

<Table IV-27> Substantive Content of Indicators

	Estimated Coefficient	z	P>z
Direct Program	0.197	1.17	0.241
Small-scale Program	-0.020	-0.11	0.911
Large-scale Program	-0.050	-0.26	0.795
Economic Program	0.101	0.45	0.652
Welfare Program	0.230	1.04	0.299
Ministry	-0.349	-1.52	0.128
Constant	2.319	9.31	0.000
Number of Samples	1717		
Rsq	0.0166		

The estimation results above show that all explanatory variables were not in a significant relationship with substantive content of indicators, indicating that the explanatory variables did not have any effect on the difference in substantive content of indicators.

The question about the connection between performance indicators and program goals is to see if performance indicators can be used to measure whether program goals have been achieved. If a program falls under any of the following category of cases, the answer to the item would be "no": 1) if the performance indicators are poorly associated with program goals; 2) if the performance indicators are focused only on the amount of input or output; 3) if the definition or estimation methods of performance indicators are not clear or reasonable; and 4) if only satisfaction indicators are presented without including out indicators that measure the quality of output and quantitative indicators.

The results in terms of the connection between performance indicators and program goals show that 28.24 percent or 923 out of 3,268 programs were given "no" from 2005 to 2011. Especially 748 programs given "no" were concentrated from 2005 to 2007, the early stage for the SABP being in place. However, after 2008 when the first three years had passed since the triennial the SABP began, the number of "no" sharply decreased to ten in 2011. The sharp decline may be mainly due to the fact that the MOSF began to examine indicators on the performance

planning stage, and discussed them with other ministries since 2009. Another important reason is the heightened education to facilitate the understanding of the importance of performance indicators. <Table IV-28> shows the estimation results of the connection between performance indicators and program goals.

<Table IV-28> Connection Between Indicators and Program Goals

	Estimated Coefficient	z	P>z
Direct Program	0.310	6.01	0.000
Small-scale Program	-0.103	-1.87	0.062
Large-scale Program	0.015	0.24	0.811
Economic Program	0.048	0.78	0.438
Welfare Program	-0.062	-0.96	0.338
Ministry	-0.143	-2.34	0.019
Constant	0.621	9.37	0.000
Number of Samples	3268		
Rsq	0.0143		

The estimation results regarding the connection between performance indicators and program goals show that there is a significant positive relationship between direct programs and the item at the 0.01 significance level, and programs carried out by ministries are negatively related to the item, and small-scale programs are negatively related to the item at the 0.1 significance level. The results indicate that in the case of direct programs, it is more likely that proper indicators are set up compared to indirect programs, and ministries are less likely to set up appropriate indicators for program goals compared to agencies and committees. In addition, though the level of significance is not high, in the case of small-scale programs, the establishment of indicators had been neglected compared to large-scale programs, which suggests that the performance of small programs is managed poorly in comparison.

6) Feasibility and Ambitiousness in Setting Target Values

The item of feasibility and ambitiousness in setting target values is designed to test the enthusiasm and ambition in the establishment of target values of performance indicators designed to accomplish program goals. If a program falls under any of the following category of cases, the answer to the questionnaire is "no", since the target values of the program cannot be regarded as reasonable or ambitious: 1) if the target values of performance indicators are set up after the program is initiated; 2) if the target values are set up at the level that can be reached without special efforts; and 3) if the target values are set up based only on the past performance without correcting problems occurring in the process of past programs.

The item of feasibility and ambitiousness in setting target values is important given that the item is associated with the evaluation item of achievement in the SABP to which the highest point is allocated. In other words, if target values are not set up in a feasible and ambitious manner, the target values are meaningless even if they are perfectly accomplished. So, if target values are not set up in a feasible and ambitious manner, only a partial score is given to the item of achievement, where the highest point is allocated, which is why the item of feasibility and ambitiousness in setting target values plays a major role in determining the rating of evaluation.

The item had always been included in the checklist from 2005 to 2011, and 70.98 percent or 2,319 out of 3,267 programs were given "no" in the questionnaire. Examination by year shows that in 2005, 2006, and 2007, the proportion of "no" was respectively 71.53 percent, 77.64 percent and 70.94 percent, all of which were over 70 percent. There was a little improvement in 2008 and 2009 with the proportion of "no," 65.56 percent and 59.25 percent, but it worsened again in 2010 and 2011 with 79.45 percent and 65.52 percent, respectively.

The proportion of "no" in indicators has sharply declined since 2008. However, the proportion of "no" decreased only a bit in the item of feasibility and ambitiousness in setting target values in 2008 and 2009, and rapidly increased again in 2010. This suggests that there was much effort toward the improvement of indicators while there was little effort toward setting proper target values. <Table IV-29> shows the estimation results

about setting proper target values.

<Table IV-29> Feasibility and Ambitiousness in Setting Target Values

	Estimated Coefficient	z	P>z
Direct Program	0.309	6.22	0.000
Small-scale Program	-0.098	-1.76	0.079
Large-scale Program	0.009	0.15	0.879
Economic Program	-0.182	-2.91	0.004
Welfare Program	-0.529	-7.29	0.000
Ministry	-0.050	-0.85	0.393
Constant	-0.482	-7.52	0.000
Number of Samples	3267		
Rsq	0.0345		

The estimation results suggest that direct, economic and welfare programs are significantly related to the item at the 0.01 level, and small-scale programs are in a significant negative relationship with the item at the 0.1 level. The coefficient of direct programs is positive, meaning that target values of direct programs were set up in a more feasible and ambitious manner than indirect programs. In addition, target values were set up in a less active way for economic and welfare programs than for general programs.

7) Monitoring

The item of monitoring is designed to check the execution process of a program, and examines whether performance data is well managed, and problems and comments from outside are properly addressed or resolved through an appropriate monitoring system. Five to ten points are allocated to this item according to year. The answer “no” is given to the item of monitoring if a program falls under any of the following category of cases: 1) if the presented information only includes performance of program management related to budget execution and the grasp of output and no effort is made to improve the quality of output; 2) if the monitoring system

does not properly deal with problems arising in the process of program management; and 3) if the National Assembly and the Board of Audit and Inspection are able to identify problems despite the ministries' statement to the contrary.

The answer "no" accounted for 12.03 percent or 402 out of 3,343 programs in the item of monitoring over the past seven years, and 70 percent or 294 programs of them were concentrated during the period of 2005-2007. The proportion of "no" was less than 10 percent in 2008 and a mere 1.44 percent in 2011, indicating that there has been a significant improvement. <Table IV-30> shows the estimation results for the item of monitoring.

<Table IV-30> Monitoring

	Estimated Coefficient	z	P>z
Direct Program	0.196	3.13	0.002
Small-scale Program	-0.080	-1.22	0.223
Large-scale Program	0.163	2.12	0.034
Economic Program	0.047	0.63	0.527
Welfare Program	0.042	0.54	0.593
Ministry	-0.221	-2.88	0.004
Constant	1.245	15.14	0.000
Number of Samples	3268		
Rsq	0.0119		

The estimation results show that there is a significant positive relationship between direct programs and the item of monitoring, and there is a significant negative relationship between ministries and the item at the 0.01 level, and large-scale programs are positively related to monitoring at the 0.05 significance level. Considering that monitoring is designed to detect and improve problems and issues in the process of program management, the results were as expected. In other words, direct, large-scale programs or programs carried out by agencies and committees performed better in the management section, compared to indirect, small-scale programs or programs under ministries as mentioned before.

8) Problem-solving

The item of problem-solving was included in the checklist from 2005 to 2009, but it has been combined with the item of monitoring since 2010.[13] Problem-solving is basically similar to monitoring in that it examines any changes in environment and whether problems are properly addressed or solved in the process of program management. Over the period of 2005 to 2009, the answer "no" was given to 12.55 percent or 307 out of 2,447 programs. <Table IV-31> shows the estimation results of problem-solving.

<Table IV-31> Problem-solving

	Estimated Coefficient	z	P>z
Direct Program	0.235	3.26	0.001
Small-scale Program	0.105	1.39	0.164
Large-scale Program	0.115	1.34	0.179
Economic Program	0.312	3.05	0.002
Welfare Program	0.048	0.54	0.592
Ministry	-0.088	-1.04	0.297
Constant	1.020	11.45	0.000
Number of Samples	2447		
Rsq	0.0130		

The estimation results demonstrate that the item of problem-solving is in a positive relationship with direct programs, and a negative relationship with ministries as seen in monitoring. However, the variable of ministries turned out not significant and there was no significant difference by size of programs. Meanwhile, there was a significant positive relationship between economic programs and the item at the 0.01 significance level.

13) After the two criteria were combined, more points have been allocated to the item of monitoring.

9) Budget Execution

The item of budget execution is designed to check if budgets are effectively managed, executed according to schedule, and used for their purposes. The answer "no" is given to this question if a program falls under any of the following category of cases: 1) if the budget is redirected without unavoidable reasons or used for purposes other than its original purpose; 2) if there is no effort made to improve poor budget execution rates; and 3) if a person disqualified for the program receives funds or if no measures are taken even though problems related to accounting occur. Since 2011, besides the answer "no," partial scores, such as "to some degree" and "to a significant degree" have been employed. More specifically, "to some degree" is given: 1) even if the execution rate is 100 percent, the budget is executed without considering the fact that a program needs to be performed at the right time (for example, heating expense aid); 2) even if the execution rate is lower than 100 percent, there has been a degree of improvement; and 3) even if funds are given to unqualified recipients or programs related to accounting occur, problems are resolved "to some degree" by improving the system. The answer of "to a significant degree" is given: 1) even if the execution rate is 100 percent, quarterly execution schedules are not complied with; and 2) even if the execution rate is lower than 100 percent, the execution rate continues to increase.

The results from 2005 to 2011 show that "no" accounts for 26.45 percent or 864 of 3,267 programs. The proportion of "no" was 38.92 percent in 2005, which is nearly 40 percent, but remained between 25 percent and 28 percent from 2006 to 2010. In 2011 when the system of partial scores was in place, the ratio of "no" sharply declined to 2.59 percent. <Table IV-32> shows the estimation results for the item of budget execution.

<Table IV-32> Budget Execution

	Estimated Coefficient	z	P>z
Direct Program	0.002	0.04	0.966
Small-scale Program	0.209	3.72	0.000
Large-scale Program	0.014	0.22	0.824
Economic Program	0.520	7.69	0.000
Welfare Program	0.257	3.8	0.000
Ministry	-0.103	-1.69	0.091
Constant	0.490	7.48	0.000
Number of Samples	3267		
Rsq	0.0235		

The estimation results demonstrate that the relationship between the variable of small-scale, economic or welfare programs and the item of budget execution is significant and positive at the 0.01 level, and ministries show a significant negative relationship with the item at the 0.1 significance level. The coefficient of small-scale programs was estimated significantly positive, in part because it is easier to execute and manage smaller budgets according to schedule. Interestingly, economic and welfare programs adhere more closely to the budget than general programs, and there was no significant difference found between direct programs and indirect programs in the item of budget execution unlike in many other items.

10) Efficiency Enhancement to Achieve Program Goals

This item is designed to assess and monitor whether a program enhances efficiency to achieve program goals. “Yes” is given if any of the followings is satisfied: 1) if line ministries/agencies cut costs by improving implementation methods and reflecting the results in budget decision the following year; 2) if ministries/agencies have received or are going to receive budgetary incentives from the MOSF under related provisions; 3) if ministries/agencies introduce internal methods to improve performance in the public sector for the first time and their actual improvement are proven; and 4) if ministries/agencies receive a positive response on the

performance from reliable outside experts, such as major media outlets, or win a prize for their program.

The intention of this item is to acknowledge extraordinary effort to cut costs. From 2005 to 2011, the proportion of programs judged as "no" was high with 64.89 percent (=2,120 / 3,267). In particular, considering that the number nears 80 percent (79.4 percent), this is the only item in the SABP that adds points to performance enhancement. <Table IV-33> presents estimation results of enhancing efficiency to achieve performance goals.

<Table IV-33> Efficiency Enhancement to Achieve Program Goals

	Estimated Coefficient	z	P>z
Direct Program	0.148	3.05	0.002
Small-scale Program	-0.168	-3.14	0.002
Large-scale Program	0.114	1.94	0.052
Economic Program	-0.263	-4.27	0.000
Welfare Program	-0.002	-0.03	0.978
Ministry	-0.192	-3.34	0.001
Constant	-0.205	-3.28	0.001
Number of Samples	3267		
Rsq	0.0163		

According to the table above, a significant positive relationship is observed between direct programs and this item of enhancing efficiency to achieve program goals at the 0.01 level, indicating that when running direct program, greater effort to boost efficiency was made than operating indirect programs. Small-scale programs show a significant negative correlation with the item at the 0.01 level, while the relationship between large-scale programs and this item is positive and significant at the 0.05 level. The results of programs in terms of their size are hardly surprising. Small-scale programs obtained poor results due to fewer opportunities to raise efficiency, with fewer options to cut costs through innovative ideas and effort. In contrast, large-scale programs showed better performance because they enjoyed the conditions that small-scale ones did not.

Economic and ministry-managed programs are in a significant negative relationship with this item at the 0.01 level.

11) Performance Goal Achievement

This item is designed to assess and monitor the extent to which line ministries/agencies achieve the intended objectives set at a planning phase, with four responses given: "no (0 point)," "to some degree (10 points)," "to a significant degree (20 points)," and "yes (30 points)." The scoring system is as follows. "Yes" is given when a program receives "yes" for the item of setting performance target values and achieves 100 percent of all the goals of performance indicators.[14] "To a significant degree" is given: 1) when a program receives "yes" for the aforementioned target values item and achieves its goals to a significant degree; and 2) in cases when, despite the program achieving 100 percent of its goals, the confirmation of achievement is difficult as the program is new or in progress, or normal verification is hindered by a problem in the execution phase. "To some degree" is given: 1) when a program receives "yes" for the target values item and achieves its goals to some degree; 2) in cases when although a program has achieved 100 percent of its goals, it presents unreliable data, or a low rate of conducting yearly assessment; and 3) when a program receives "no" for the target values item and achieves 100 percent of goals. "No" is given when none of the above mentioned conditions are satisfied.

This item accounts for the largest share of SABP item scores and is related to setting planning-stage goals. <Table IV-34> shows yearly averages of this item.

14) Even if a rate of goal achievement exceeds 100 percent, the final rate is limited to 100 percent.

<Table IV-34> Yearly Averages of Achieving Performance Goals

Year	Average
2005	12.65
2006	13.58
2007	15.51
2008	15.90
2009	20.87
2010	11.97
2011	13.03

As presented above, the yearly averages had continued to be on the rise from 2005 to 2009 but showed a rapid decline since 2010. The steady increase between 2005 and 2009 indicates that performance management had been settled down. The drop in the average ratings since 2010 is not because performance management was neglected, but because of a change in the scoring system which added since 2010 questions to check the connectivity with upper performance goals. In the changed scoring system, if upper performance goals were not met, unit projects under the same upper performance goal underwent a five-point cut in scores.[15)]

<Table IV-35> presents estimation results of achieving performance goals. Unlike other evaluation items, it allowed partial scores and used OLS method, not a probit model, because its scores had obvious meanings.

15) Another possible reason may be the stricter inspection of setting and achieving goals since 2010, when data reliability of performance planning and performance reports came under investigation by the Board of Audit and Inspection. But this is presented as a mere possibility because it is not verified with the data used in the analysis of this subsection.

<Table IV-35> Achieving Performance Goals 1 (OLS)

	Estimated Coefficient	t	P>t
Direct Program	1.902	6.93	0.000
Small-scale Program	-0.584	-1.95	0.051
Large-scale Program	0.210	0.63	0.527
Economic Program	-0.079	-0.23	0.818
Welfare Program	-1.700	-4.70	0.000
Ministry	-1.412	-4.35	0.000
Year Dummy	yes		
Constant	13.357	30.50	0.000
Number of Samples	3268		
Adj-Rsq	0.1410		

The above table shows that the relationship between indirect programs and this item of achieving performance goals is positive and significant at the 0.01 level. Small-scale programs show a significant negative relationship with this item at the 0.05 level. Both welfare and ministry-managed programs are positively related with the item, showing a significant correlation at the 0.01 level. However, in order to obtain more precise estimates of achieving performance goals, it is required to consider the link between setting an ambitious target at a planning stage and their actual results because they are related, as previously noted, with the scoring system of goal achievement. <Table IV-36> illustrates new results of the programs which received "yes" for the 6th item of feasibility and ambitiousness in setting target values.

<Table IV-36> Achieving Performance Goals 2

	Estimated Coefficient	t	P>t
Direct Program	1.022	1.83	0.067
Small-scale Program	-0.317	-0.50	0.617
Large-scale Program	0.483	0.70	0.482
Economic Program	1.362	1.77	0.076
Welfare Program	-1.425	-1.47	0.143
Ministry	-2.424	-3.73	0.000
Year Dummy	yes		
Constant	20.851	22.120	0.000
Number of Samples	948		
Adj-Rsq	0.0754		

According to the above table, only ministry-managed programs are in a significant negative correlation with the item of achieving performance goals at the 0.01 level. Both direct and economic programs are positively related with it, showing a significant relationship at the 0.1 level. Considering that the relationship between ministry-managed programs and this item is negative and significant at the 0.01 level, even if ministry-managed programs set a feasible and ambitious goal, they achieve lower results than agency/committee-managed programs, thereby highlighting the need for ministries to make more efforts to achieve their performance goals.

12) Program Evaluation Implementation

This item is designed to examine whether efficiency assessment of a program was conducted by means of an objective and comprehensive analysis based on reliable data. When all of the following conditions are fulfilled, a program is given "yes" on this item: 1) conducting an in-depth evaluation by utilizing verifiable data and specialized analytical methods; 2) making a comprehensive assessment on the whole programs; and 3) having performance evaluation objectively conducted by an independent organization.

On the other hand, the following cases receive "no": 1) making an internal evaluation; 2) assessing only subprograms or the processes of them; and 3) using unreliable data or inappropriate evaluation technique. In addition, assessing performance indicators by using a checklist does not constitute an in-depth evaluation required by this item, nor does evaluation by the National Assembly or the Board of Audit and Inspection which focus on the problems arising in the process of implementing a program. One typical example of program evaluation is an in-depth evaluation conducted under the National Finance Act.

To be sure, it is true that small-scale programs are at a disadvantage because they cannot afford time and cost required to conduct program assessment.[16] Nevertheless, this item was required due to the benefits of conducting triennial program assessments in order for greater understanding and enhanced inefficiency of the program. Since 2011, however, programs with a budget of less than one billion won have been exempted from program assessment as many continued to pose objections that small-scale programs are at a disadvantage.

This item was sometimes employed to check only whether or not program evaluation was implemented or sometimes whether recommendations made from evaluation were reflected, thereby improving the program along with questions to ask whether or not program evaluation was implemented. This subsection estimates the results of conducting program assessment with the data only related to the former case. <Table IV-37> illustrates the estimated results.

16) It is assumed that the need to conduct an actual assessment was not felt because the range and efficiency of small-scale programs were easier to discern compared to large-scale ones.

<Table IV-37> Program Assessment Implementation

	Estimated Coefficient	z	P>z
Direct Program	-0.223	-4.59	0.000
Small-scale Program	-0.324	-6.14	0.000
Large-scale Program	0.180	3.02	0.003
Economic Program	-0.147	-2.45	0.014
Welfare Program	0.103	1.59	0.111
Ministry	-0.140	-2.43	0.015
Constant	0.465	7.41	0.000
Number of Samples	3194		
Rsq	0.0259		

According to the above table, as expected, a significant negative relationship is observed between small-scale programs and this item of implementing program assessment at the 0.01 level, whereas large-scale programs are in a significant positive correlation with this item at the same significance level. As for direct programs, the relationship with this item is significant negative at the 0.01 level, with economic programs in a significant positive correlation at the same level. There is no statistical significance in the relationship between welfare programs and this item. Ministry-managed programs are in a significant negative correlation with the item at the 0.05 level, showing that ministries were less likely to conduct program assessment than agencies and committees.

13) Feedback

This item is designed to confirm whether ministries/agencies reflect the results of self-assessment or program evaluation in improving systems. If a program falls under any of the following categories, it is given "no": 1) if there are no records of improving systems by taking complementary measures based on program evaluation, or it has failed to work out a plan to improve systems; and 2) if there are no records of reflecting recommendations by the MOSF through self-assessment and an in-depth

evaluation.[17)]

In this item, the proportion of programs judged as "no" was 43.7 percent and 44.3 percent in 2005 and 2006, respectively when the item was introduced. However, it plunged to 20.9 percent in 2007, declining to 18.0 percent in 2008. Since 2010, almost all programs have strived to resolve problems, which were revealed through program evaluation or an in-depth assessment, with the proportion of 1.8 percent in 2010 and 0.6 percent in 2011. The estimation results on feedback are presented in <Table IV-38>.

<Table IV-38> Feedback

	Estimated Coefficient	z	P>z
Direct Programs	-0.137	-2.60	0.009
Small-scale Programs	-0.059	-1.02	0.308
Large-scale Programs	0.151	2.29	0.022
Economic Programs	0.235	3.46	0.001
Welfare Programs	0.234	3.31	0.001
Ministry	-0.340	-5.25	0.000
Constant	0.968	13.75	0.000
Number of Samples	3219		
Rsq	0.0155		

According to the results, only the variable of program size is not in a significant relationship with this item of feedback at the 0.01 level, with other variables in a significance correlation with the item at the same level. In terms of the variable of program size, large-scale programs are in a significant positive correlation with the item at the 0.05 level. The results revealed the notable fact that direct programs provide less feedback than indirect programs. As previously noted, direct programs outperform indirect programs in almost all items when it comes to performance management. But the case is not true for only this item of feedback, indicating that direct programs need to pay greater attention to feedback.

17) A "no" is always given for programs with no records of resolving problems identified in previous evaluations.

D. Programs Rated as Ineffective

So far, the discussion has focused on the characteristics that affect total score, scores by sector, and scores by item. Although scores also hold significance in the SABP, being rated as ineffective or very ineffective has a more direct impact on the operation of a program. Therefore, in this subsection, only programs rated as ineffective with the score less than 60 were selected to discuss the characteristics of a program that lead to ineffective ratings.

<Table IV-4> presents programs evaluated as ineffective and very ineffective by year. This subsection employs a probit estimation with ineffective and very ineffective ratings denoted as "1," and other ratings as "0." The estimation results are shown in <Table IV-39>.

<Table IV-39> Ineffective Programs

	Estimated Coefficient	z	P>z
Direct Program	-0.256	-4.440	0.000
Small-scale Program	0.132	2.190	0.029
Large-scale Program	-0.231	-3.320	0.001
Economic Program	0.161	2.400	0.016
Welfare Program	0.121	1.720	0.086
Ministry	0.227	3.230	0.001
Constant	-1.088	-14.450	0.000
Number of Samples	3343		
Rsq	0.0253		

The results are not much different from estimation results when scores are used as dependent variable. Direct programs are in a significant negative relationship with the item of earning ineffective and very ineffective ratings at the 0.01 level, indicating that direct programs received less ineffective ratings. Small-scale programs received more ineffective ratings than large-scale programs, economic or welfare programs than general ones. Ministry-managed programs earned more ineffective ratings than agency/committee-managed ones.

6 The Analysis of Welfare Programs

A. Current State of Analysis Target Program

In this subsection, further analysis on welfare programs was conducted in order to examine whether additional issues can be deduced in consideration of increasing demand on the welfare budget. Among many welfare programs, the emphasis was placed on local-government and private-sector subsidy programs and a comparison between the two types of programs was made. Although a comparison of welfare programs by support type (vouchers / cash / goods) was attempted, it was impossible to derive a significant analysis because there was a large gap in the numbers of each type program and certain programs were rarely conducted.

To select welfare-focused programs, contents and purpose of an individual program were examined, rather than applying the standard of line ministries / agencies. As a result, 100 programs were confirmed to have conducted the SABP between 2009 and 2012. The time period was limited to 2009-2012 because other programs were not available for figuring out the details.

In terms of budget execution, 13 percent out of 100 programs were conducted through direct execution, investment, SOC, equipment procurement, while 87 percent was indirectly made through private-sector subsidy, local-government support, loans, contributions and investments. For each year that the SABP was conducted, the number of programs selected for analysis was 16 from 2009, 20 from 2010, 22 from 2011, and 42 from 2012 with programs from 2012 accounting for the largest part.

By operating organizations, 88 programs were ministry-managed, with 12 programs agency-managed. The largest number of programs was conducted by the Ministry of Health and Welfare with 43 programs, followed by the Ministry of Employment and Labor with 14 programs. In terms of financial resources, 49 programs were supported by funds, and 51 by general accounts. As for spending type, 86 programs were discretionary spending programs, while nine were mandatory spending programs[18)]

18) Mandatory spending programs include the Infant Care Subsidy, Disability Allowance, Self-

with five of other types. By aid type, 73 programs were supported by cash, followed by goods and vouchers. By the field of programs, the largest number of programs was related with health and medical services, followed by job creation, miscellaneous services, residence and energy, education and scholarship, and child care.

By the year of program being in place, the oldest program was conducted in 1953, while the latest was in 2011, with the average year being 1999 (standard deviation=11.7).

The average budget of the assessed programs was 331.8 billion won (standard deviation=9937.11). The lowest budget was 400 million won for Reverse Mortgage Loans for Low-earners' Old Age Pensions by the Financial Services Commission, while the highest was 6,800 billion won for the House Purchase & Rent Subsidy by the Ministry of Land, Transport and Maritime Affairs.

sufficiency Program, Basic Senior Pensions, Start-up Assistance for Long-term Unemployed, Basic Livelihood Security Subsidy by the Ministry of Health and Welfare, and Single-parent Family Support by the Ministry of Gender Equality and Family.

<Table IV-40> Characteristics of Analyzed Welfare Programs

Variable		Frequency	%
Assessment Year	2009	16	16.0
	2010	20	20.0
	2011	22	22.0
	2012	42	42.0
Operating Organization	Agency	12	12.0
	Ministry	88	88.0
Financial Resources	Fund	49	49.0
	General Accounts	51	51.0
Spending Type	Discretionary	86	86.0
	Mandatory	9	9.0
	Miscellaneous	5	5.0
Program Type	Direct	13	13.0
	Indirect	87	87.0
Support Type	Voucher	6	6.0
	Cash	73	73.0
	Goods	21	21.0
Program Field	Education & Scholarship	11	11.0
	Health & Medical Services	27	27.0
	Child Care	10	10.0
	Job Creation	23	23.0
	Residence & Energy	13	13.0
	Miscellaneous	16	16.0
Total		100	100.0

B. The Analysis of External Comment Type

In order to compile external evaluation of the relevant programs, this study has used data from the Board of Audit and Inspection, and the Special Committee of Budget and Accounts of the National Assembly.

Out of a total of 100 budgetary programs related with welfare spending, 88 programs received comments from outside experts in terms of systems, budget, management and 12 programs received no comment, with welfare programs having an average of 1.98 comments (standard deviation=0.995).

<Table IV-41> Extend Comments by Program Type

(Unit: number of frequency, %)

		Frequency of External Comment					
		0	1	2	3	4	Total
Assessment Year	2009	2	2	7	5	0	16
		16.7	18.2	14.9	18.5	0.0	16.0
	2010	4	3	10	2	1	20
		33.3	27.3	21.3	7.4	33.3	20.0
	2011	1	3	10	7	1	22
		8.3	27.3	21.3	25.9	33.3	22.0
	2012	5	3	20	13	1	42
		41.7	27.3	42.6	48.1	33.3	42.0
Operating Organization	Agency	3	1	4	4	0	12
		25.0	9.1	8.5	14.8	0.0	12.0
	Ministry	9	10	43	23	3	88
		75.0	90.9	91.5	85.2	100.0	88.0
Financial Resources	Fund	7	3	27	11	1	49
		58.3	27.3	57.4	40.7	33.3	49.0
	General accounts	5	8	20	16	2	51
		41.7	72.7	42.6	59.3	66.7	51.0

<Table IV-41> Continue

(Unit: number of frequency, %)

		Frequency of External Comment					
		0	1	2	3	4	Total
Spending Type	Discretionary	10	9	46	20	1	86
		83.3	81.8	97.9	74.1	33.3	86.0
	Mandatory	1	0	0	6	2	9
		8.3	0.0	0.0	22.2	66.7	9.0
	Miscellaneous	1	2	1	1	0	5
		8.3	18.2	2.1	3.7	0.0	5.0
Program Type	Direct	2	3	5	3	0	13
		16.7	27.3	10.6	11.1	0.0	13.0
	Indirect	10	8	42	24	3	87
		83.3	72.7	89.4	88.9	100.0	87.0
Support Type	Vouchers	0	1	1	4	0	6
		0.0	9.1	2.1	14.8	0.0	6.0
	Cash	9	7	36	18	3	73
		75.0	63.6	76.6	66.7	100.0	73.0
	Goods	3	3	10	5	0	21
		25.0	27.3	21.3	18.5	0.0	21.0
Program Field	Education & Scholarship	1	0	7	3	0	11
		8.3	0.0	14.9	11.1	0.0	11.0
	Health & Medical Services	5	7	11	4	0	27
		41.7	63.6	23.4	14.8	0.0	27.0
	Child Care	0	1	6	3	0	10
		0.0	9.1	12.8	11.1	0.0	10.0
	Job Creation	4	1	11	7	0	23
		33.3	9.1	23.4	25.9	0.0	23.0
	Residence & Energy	1	1	7	3	1	13
		8.3	9.1	14.9	11.1	33.3	13.0
	Miscellaneous	1	1	5	7	2	16
		8.3	9.1	10.6	25.9	66.7	16.0
Total		12	11	47	27	3	100
		100.0	100.0	100.0	100.0	100.0	100.0

Types of comments from outside are organized by the characteristics of welfare programs as in the following table.

<Table IV-42> Types of Extend Comments

(Unit: number of frequency, %)

		Systemic Issues	Budget Issues	Management Issues	Miscellaneous	Total
Assessment Year	2009 (16)	11	10	10	0	14
		78.6	71.4	71.4	0	15.9
		68.8	62.5	62.5	0	87.5
	2010 (20)	14	5	11	3	16
		87.5	31.3	68.8	18.8	18.2
		70	25	55	15	80
	2011 (22)	17	12	17	2	21
		81.0	57.1	81.0	9.5	23.9
		77.3	54.5	77.3	9.1	95.5
	2012 (42)	32	21	31	2	37
		86.5	56.8	83.8	5.4	42.0
		76.2	50.0	73.8	4.8	88.1
Program Type	Direct (13)	10	5	7	0	11
		90.9	45.5	63.6	0	12.5
		76.9	38.5	53.8	0.0	84.6
	Indirect (87)	64	43	62	7	77
		83.1	55.8	80.5	9.1	87.5
		73.6	49.4	71.3	8.0	88.5
Spending Type	Discretionary (86)	63	38	59	5	76
		82.9	50	77.6	6.6	86.4
		73.3	44.2	68.6	5.8	88.4
	Mandatory (9)	8	8	8	2	8
		100.0	100.0	100.0	25.0	9.1
		88.9	88.9	88.9	22.2	88.9
	Miscellaneous (5)	3	2	2	0	4
		75.0	50.0	50.0	0.0	4.5
		60.0	40.0	40.0	0.0	80.0

<Table IV-42> Continue

(Unit: number of frequency, %)

		Systemic Issues	Budget Issues	Management Issues	Miscellaneous	Total
Financial Resources	Fund (49)	30	20	32	4	42
		71.4	47.6	76.2	9.5	47.7
		61.2	40.8	65.3	8.2	85.7
	General accounts (51)	36	28	37	3	46
		78.3	60.9	80.4	6.5	52.3
		70.6	54.9	72.5	5.9	90.2
Operating Organization	Agency (12)	8	6	7	0	9
		88.9	66.7	77.8	0	10.2
		66.7	50.0	58.3	0	75
	Ministry (88)	66	42	62	7	79
		83.5	53.2	78.5	8.9	89.8
		75.0	47.7	70.5	8.0	89.8
Support Type	Vouchers (6)	6	4	5	0	6
		100.0	66.7	83.3	0.0	6.8
		100.0	66.7	83.3	0.0	6.8
	Cash (73)	54	35	51	5	64
		84.4	54.7	79.7	7.8	72.7
		74.0	47.9	69.9	6.8	87.7
	Goods (21)	14	9	13	2	18
		77.8	50.0	72.2	11.1	20.5
		66.7	42.9	61.9	9.5	85.7

<Table IV-42> Continue

(Unit: number of frequency, %)

		Systemic Issues	Budget Issues	Management Issues	Miscellaneous	Total
Program Field	Education & Scholarship (11)	8	5	10	0	10
		80.0	50.0	100.0	0.0	11.4
		72.7	45.5	90.9	0.0	90.9
	Health & Medical Services (27)	18	11	12	0	22
		81.8	50.0	54.5	0.0	25.0
		66.7	40.7	44.4	0.0	81.5
	Child Care (10)	10	4	4	4	10
		100.0	40.0	40.0	40.0	11.4
		100.0	40.0	40.0	40.0	100.0
	Job Creation (23)	18	8	18	0	19
		94.7	42.1	94.7	0	21.6
		78.3	34.8	78.3	0.0	82.6
	Residence & Energy (13)	7	10	10	1	12
		58.3	83.3	83.3	8.3	13.6
		53.8	76.9	76.9	7.7	92.3
	Miscellaneous (16)	13	10	15	2	15
		86.7	66.7	100.0	13.3	17.0
		81.3	62.5	93.8	12.5	93.8
Total		74	48	69	7	88
		84.1	54.5	78.4	8.0	100

The state of external evaluation of programs organized by program type is as follows.

Out of 13 directly managed programs, 11 received external comment. System problems received the most comments, followed by management and budget issues.

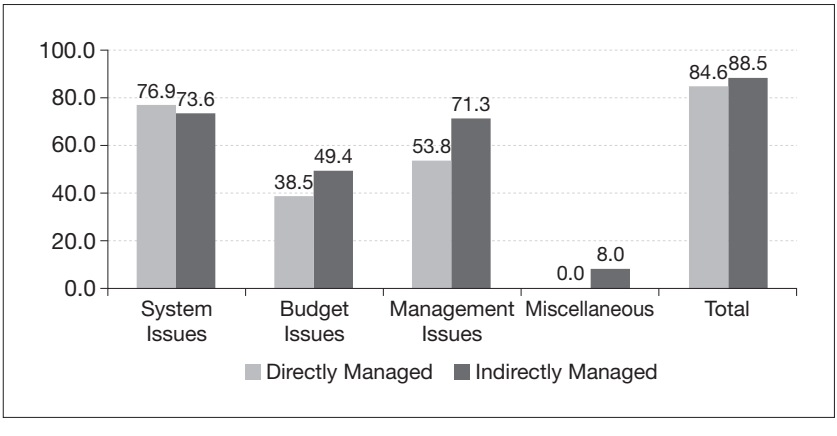
As for indirect programs, 77 out of 87 programs were commented by external experts, with the frequency of comment in the same order as direct programs.

As a whole, indirect programs had a high rate of earning comments, except for in the subsection of system problems.

<Table IV-43> Types of Extend Comments by Program Management Type

	Systemic Issues	Budget Issues	Management Issues	Miscellaneous	Total
Directly Managed (13)	10	5	7	0	11
	76.9	38.5	53.8	0.0	84.6
Indirectly Managed (87)	64	43	62	7	77
	73.6	49.4	71.3	8.0	88.5
Total	74	48	69	7	88
	84.1	54.5	78.4	8.0	100

[Figure IV-7] Types of Extend Comments by Program Management Type



C. The Analysis of SABP Ratings by Characteristics of Programs

Fund programs were less likely to receive ineffective ratings than general accounting programs. A mere five out of 49 fund programs were rated as ineffective, while 10 out of 51 general accounting programs earned ineffective ratings. Two out of 12 agency-managed programs, or 16.7 percent, were rated as ineffective, while 13 out of 88 ministry-managed programs, or 14.8 percent, earned ineffective ratings.

In terms of spending type, 12 out of 86 discretionary spending programs were evaluated as ineffective, while three out of nine mandatory

spending programs earned ineffective ratings. While 33.3 percent of mandatory spending programs were rated as ineffective, discretionary spending programs had a low proportion of ineffective ratings with 13.9 percent. As for ratings by program type, three out of 13 direct programs (23.1 percent) received the ineffective rating, with 12 out of 87 indirect programs (13.8 percent) rated as ineffective.

When it comes to support type, a total of six voucher programs had no ineffective ratings, 12 out of 73 cash-supported programs (16.4 percent) and three out of 21 goods-supported programs (14.3 percent) earned ineffective ratings. As for program field, job creation programs had a notably high proportion of ineffective ratings.

<Table IV-44> Assessment Results by Program Type

		Ineffective	Adequate	Effective	Total
Assessment Year	2009	1	14	1	16
		6.7	17.9	14.3	16.0
	2010	2	18	0	20
		13.3	23.1	0.0	20.0
	2011	5	13	4	22
		33.3	16.7	57.1	22.0
	2012	7	33	2	42
		46.7	42.3	28.6	42.0
Financial Resources	Fund	5	42	2	49
		33.3	53.8	28.6	49.0
	General Accounting	10	36	5	51
		66.7	46.2	71.4	51.0
Operating Organization	Agency	2	9	1	12
		13.3	11.5	14.3	12.0
	Ministry	13	69	6	88
		86.7	88.5	85.7	88.0

<Table IV-44> Continue

		Ineffective	Adequate	Effective	Total
Spending Type	Discretionary	12	67	7	86
		80.0	85.9	100.0	86.0
	Mandatory	3	6	0	9
		20.0	7.7	0.0	9.0
	Miscellaneous	0	5	0	5
		0.0	6.4	0.0	5.0
Program Type	Direct	3	10	0	13
		20.0	12.8	0.0	13.0
	Indirect	12	68	7	87
		80.0	87.2	100.0	87.0
Support Type	Voucher	0	5	1	6
		0.0	6.4	14.3	6.0
	Cash	12	56	5	73
		80.0	71.8	71.4	73.0
	Goods	3	17	1	21
		20.0	21.8	14.3	21.0
Program Field	Education & Scholarship	2	7	2	11
		13.3	9.0	28.6	11.0
	Health & Medical Services	2	25	0	27
		13.3	32.1	0.0	27.0
	Child Care	1	9	0	10
		6.7	11.5	0.0	10.0
	Job Creation	6	14	3	23
		40.0	17.9	42.9	23.0
	Residence & Energy	1	12	0	13
		6.7	15.4	0.0	13.0
	Miscellaneous	3	11	2	16
		20.0	14.1	28.6	16.0
Total		15	78	7	100
		100	100	100	100

D. The Analysis of Variables' Effect on SABP Total Scores

First of all, the effect of major evaluation items on SABP total scores was analyzed. Model 1 shows the effect on SABP total scores imposed by assessment items such as program design, performance planning, budget execution, monitoring, and program evaluation.

Further analysis shows that three items had significant effects on the scores: 1) performance planning which evaluates the adequacy of performance indicators and target values; 2) budget execution which assesses the appropriateness of budge implementation; and 3) program evaluation which rates the efficiency of the program through comprehensive and objective program evaluation.

In particular, performance planning had the most significant effect on the total scores, followed by budget execution and program evaluation, while program design and monitoring did not have a significant effect.

Model 2 illustrates how variables of program characteristics, such as spending type, program type, support type, operating organization, evaluation year, program-starting year, had effect on SABP total scores.

According to the results, mandatory spending type, operating organization, and budget size had a significant effect on the total scores. In other words, when a program was mandatory spending-related and ministry-managed with a smaller budget, the program had lower SABP total scores. In particular, budget size was the most definitive factor, followed by mandatory spending type and operating organization. In contrast, support type, program type (direct programs / indirect programs), program-starting year, and evaluation year had no significant effect.

Mandatory spending programs had more ineffective ratings than discretionary programs in terms of performance planning and budget execution, although they received higher ratings in monitoring and program evaluation. Therefore, it is necessary to make more accurate budget estimates for mandatory spending programs. In addition, it must be ascertained whether such programs are taking advantage of the mandatory expenditure to justify their passivity in performance planning, since mandatory programs rarely face cost cuts in accordance with evaluation results.

Welfare programs had poorer performance results than general administration or economic programs. According to the analysis mentioned above, the main contributing factor is that most welfare programs were managed by related ministries and mandatory spending programs exhibited low performance.

<Table IV-45> Effects on Total Score of Assessment

		Non-Standardized Regression Coefficient		Standardized Regression Coefficient	t	Sig.
		B	Std. Error	Beta		
Model 1	(Constant)	35.72	4.26		8.38	0.00
	Program Design	-5.23	4.27	-0.10	-1.23	0.22
	Performance Planning	28.40	3.18	0.77	8.94	0.00
	Budgeting	10.84	2.03	0.38	5.34	0.00
	Monitoring	6.34	3.02	0.16	2.10	0.04
	Program Assessment	5.60	1.99	0.21	2.82	0.01
	R_5=.513, F=21.664***					
Model 2	(Constant)	84.51	179.83		0.47	0.64
	Miscellaneous	0.61	4.72	0.01	0.13	0.90
	Mandatory Spending	-9.65	4.15	-0.29	-2.32	0.02
	Vouchers	1.92	4.21	0.05	0.46	0.65
	Goods	-2.23	2.61	-0.09	-0.86	0.39
	Financial Resources	0.33	2.21	0.02	0.15	0.88
	Operating Organization	-6.66	3.31	-0.22	-2.01	0.05
	Program Type	3.42	3.44	0.12	0.99	0.32
	Budget Size	3.08	1.22	0.29	2.53	0.01
	Assessment Year	-0.69	1.00	-0.08	-0.69	0.49
	Program-starting Year	-0.01	0.09	-0.01	-0.09	0.93
	R_2=.127, F=1.291					

E. The Analysis of SABP Results by Program Type

The following is comparison by program phases among 100 programs subjected to analysis, which have been grouped by financial resources, management organization, program type, support type, and program field.

In terms of financial resources, fund programs outperformed general programs. In particular, they show better performance at a planning phase including program and performance planning.

In terms of operating organization, agency-managed programs outperformed ministry-managed programs. Before examining program achievements, it should be noted that only 12 programs were agency-managed programs, making significant analysis difficult. It was expected that agencies would surpass ministries in actual performance, but evaluation results of budget execution show the opposite trend. However, agencies achieved better results than ministries in program and performance planning, suggesting that agency-managed programs had less overlaps or similarities and had clearer performance indicators and target values.

In terms of spending type, mandatory spending programs showed poorer performance. In other analyses with controlled variables, mandatory spending programs earned ineffective ratings, indicating that they received low performance results regardless of financial resources, program type, budget size, operating organization, support type, and duration period. In particular, they showed notably low results in performance planning and budget execution. It can be inferred that ministries / agencies conducting mandatory programs are less aggressive in setting performance indicators and target values and put less effort in managing programs on the grounds that mandatory spending programs are unlikely to have their cost cut. In dealing with financial resources, they also showed poor performance results because of inaccurate budget estimation and a lack of effort in budget execution.

In terms of support type, voucher-supported programs showed better performance. However, it should be noted that the number of the programs was limited to six. It is assumed that voucher programs achieved

better results because they were newly introduced, driving performers to be aggressive in program design or management.

In terms of program field, child care and job creation programs received results below the average. Child care programs had below-average results in performance planning and goal achievement, while job creation programs in budget execution.

<Table IV-46> Assessment Results by Program Type

			Total Score	Planning phase	Program Planning	Performance Planning	Management Phase	Budget Execution	Monitoring	Result Phase	Program Assessment	N
Assessment Year	2009	Average	65.99	0.78	0.65	0.56	0.77	0.60	0.72	0.54	0.53	16
		Standard Deviation	5.71	0.13	0.16	0.25	0.15	0.30	0.18	0.12	0.49	
	2010	Average	63.20	0.73	0.73	0.49	0.93	0.71	0.77	0.46	0.61	20
		Standard Deviation	2.36	0.08	0.13	0.04	0.13	0.52	0.10	0.06	0.31	
	2011	Average	65.94	0.78	0.86	0.66	0.67	0.59	0.73	0.61	0.57	22
		Standard Deviation	12.37	0.16	0.18	0.28	0.18	0.27	0.26	0.18	0.42	
	2012	Average	63.84	0.87	0.98	0.77	0.55	0.58	0.70	0.60	0.54	42
		Standard Deviation	11.42	0.17	0.11	0.29	0.18	0.29	0.31	0.16	0.34	
	Total	Average	64.52	0.81	0.85	0.66	0.69	0.61	0.72	0.56	0.56	100
		Standard Deviation	9.69	0.16	0.19	0.27	0.22	0.34	0.25	0.15	0.37	
Financial Resources	Fund	Average	64.66	0.85	0.89	0.73	0.64	0.61	0.70	0.57	0.56	49
		Standard Deviation	10.77	0.17	0.19	0.28	0.22	0.29	0.29	0.15	0.37	
	General Accounts	Average	64.38	0.77	0.81	0.58	0.73	0.62	0.75	0.55	0.56	51
		Standard Deviation	8.62	0.13	0.18	0.24	0.21	0.39	0.21	0.16	0.39	
	Total	Average	64.52	0.81	0.85	0.66	0.69	0.61	0.72	0.56	0.56	100
		Standard Deviation	9.69	0.16	0.19	0.27	0.22	0.34	0.25	0.15	0.37	

<Table IV-46> Continue

			Total Score	Planning phase	Program Planning	Performance Planning	Management Phase	Budget Execution	Monitoring	Result Phase	Program Assessment	N
Operating Organization	Agency	Average	68.30	0.92	0.97	0.87	0.58	0.49	0.83	0.64	0.38	12
		Standard Deviation	9.56	0.11	0.10	0.22	0.15	0.33	0.34	0.11	0.38	
	Ministry	Average	64.00	0.79	0.83	0.63	0.70	0.63	0.71	0.55	0.58	88
		Standard Deviation	9.64	0.16	0.19	0.27	0.22	0.34	0.24	0.15	0.37	
	Total	Average	64.52	0.81	0.85	0.66	0.69	0.61	0.72	0.56	0.56	100
		Standard Deviation	9.69	0.16	0.19	0.27	0.22	0.34	0.25	0.15	0.37	
Spending Type	Discretionary	Average	64.97	0.83	0.85	0.69	0.67	0.62	0.70	0.57	0.54	86
		Standard Deviation	10.22	0.15	0.19	0.27	0.21	0.28	0.26	0.16	0.38	
	Mandatory	Average	59.82	0.69	0.85	0.44	0.71	0.36	0.89	0.51	0.69	9
		Standard Deviation	4.62	0.10	0.19	0.17	0.21	0.30	0.11	0.08	0.28	
	Miscellaneous	Average	65.20	0.73	0.80	0.57	0.88	0.91	0.88	0.53	0.56	5
		Standard Deviation	2.28	0.21	0.21	0.25	0.28	0.92	0.11	0.10	0.40	
	Total	Average	64.52	0.81	0.85	0.66	0.69	0.61	0.72	0.56	0.56	100
		Standard Deviation	9.69	0.16	0.19	0.27	0.22	0.34	0.25	0.15	0.37	

<Table IV-46> Continue

			Total Score	Planning phase	Program Planning	Performance Planning	Management Phase	Budget Execution	Monitoring	Result Phase	Program Assessment	N
Program Type	Direct	Average	64.05	0.89	1.00	0.78	0.51	0.54	0.65	0.62	0.42	13
		Standard Deviation	8.97	0.12	0.00	0.24	0.17	0.29	0.32	0.12	0.40	
	Indirect	Average	64.59	0.80	0.83	0.64	0.71	0.62	0.74	0.55	0.58	87
		Standard Deviation	9.84	0.16	0.19	0.27	0.21	0.35	0.24	0.15	0.37	
	Total	Average	64.52	0.81	0.85	0.66	0.69	0.61	0.72	0.56	0.56	100
		Standard Deviation	9.69	0.16	0.19	0.27	0.22	0.34	0.25	0.15	0.37	
Support Type	Vouchers	Average	67.30	0.88	0.86	0.75	0.73	0.66	0.82	0.56	0.63	6
		Standard Deviation	6.83	0.14	0.16	0.27	0.07	0.34	0.18	0.13	0.38	
	Cash	Average	64.60	0.82	0.87	0.67	0.65	0.58	0.73	0.58	0.55	73
		Standard Deviation	9.28	0.16	0.18	0.29	0.21	0.29	0.26	0.15	0.37	
	Goods	Average	63.44	0.76	0.76	0.59	0.80	0.70	0.68	0.51	0.56	21
		Standard Deviation	11.80	0.15	0.20	0.21	0.24	0.48	0.22	0.16	0.39	
	Total	Average	64.52	0.81	0.85	0.66	0.69	0.61	0.72	0.56	0.56	100
		Standard Deviation	9.69	0.16	0.19	0.27	0.22	0.34	0.25	0.15	0.37	

<Table IV-46> Continue

			Total Score	Planning phase	Program Planning	Performance Planning	Management Phase	Budget Execution	Monitoring	Result Phase	Program Assessment	N
Program Type	Education & Scholarship	Average	66.91	0.88	0.87	0.77	0.55	0.36	0.63	0.64	0.49	11
		Standard Deviation	9.59	0.18	0.23	0.34	0.21	0.30	0.28	0.14	0.32	
	Health & Medical Services	Average	64.96	0.79	0.82	0.61	0.72	0.67	0.75	0.56	0.52	27
		Standard Deviation	7.03	0.17	0.19	0.28	0.18	0.25	0.20	0.14	0.41	
	Child Care	Average	63.20	0.79	0.84	0.59	0.78	0.63	0.86	0.49	0.66	10
		Standard Deviation	4.11	0.09	0.21	0.18	0.20	0.30	0.16	0.06	0.34	
	Job Creation	Average	63.60	0.81	0.91	0.68	0.61	0.57	0.70	0.58	0.63	23
		Standard Deviation	12.51	0.16	0.14	0.28	0.22	0.34	0.29	0.16	0.35	
	Residence & Energy	Average	63.78	0.85	0.87	0.73	0.65	0.61	0.62	0.56	0.47	13
		Standard Deviation	13.61	0.16	0.19	0.26	0.27	0.26	0.36	0.19	0.39	
	Miscellaneous	Average	64.87	0.78	0.78	0.61	0.82	0.72	0.78	0.52	0.56	16
		Standard Deviation	8.76	0.15	0.20	0.23	0.16	0.52	0.12	0.16	0.41	
	Total	Average	64.52	0.81	0.85	0.66	0.69	0.61	0.72	0.56	0.56	100
		Standard Deviation	9.69	0.16	0.19	0.27	0.22	0.34	0.25	0.15	0.37	

V

Summary and Conclusion

1 Summary of Analysis Results

This paper attempted to analyze performance by program type through the use of SABP data collected between 2005 and 2011. The following is a summation of analysis conducted on all programs, with control measures imposed on other variables that may affect evaluation results.

- Direct programs with direct government services and management outperformed indirect programs under which the government outsources services to other organizations.
- Welfare programs showed poorer performance than other programs.
- Program performance showed a positive correlation to the budge size.
- Agency-managed programs received relatively favorable ratings.

More detailed results are illustrated as in <Table V-1>.

First, indirect program received poorer ratings than direct programs in program design, program management, and results. In almost all specific evaluation standards, indirect programs were outperformed by direct

programs. Contrary to expectations, results for indirect programs failed to show significant differences against direct programs in budget execution, and received favorable ratings in terms of program evaluation and feedback. It was also unexpected that the budget execution revealed as many problems for direct programs as indirect programs, which will be discussed in the next subsection. In terms of implementing program evaluation and feedback of assessment results, indirect programs excelled over direct programs presumably because line ministries/agencies of indirect programs were under pressure to manage and evaluate their programs and faced many elements of risk in the management process. However, indirect programs earned ineffective ratings in most other items such as clarity of purpose, redundancy, adequacy of the execution method, adequacy of performance indicators and target values, monitoring and problem solving, efficiency enhancement, and performance goal achievement. Such shortcomings hint at the risk facing indirect programs in terms of program planning and management, and achieving performance goals, resulting in poor performance results. More detailed case analysis will be discussed in the next subsection. The number of indirect programs is expected to be on the rise for the following reasons: 1) the number of welfare-related programs is increasing; 2) services are becoming more specific; and 3) the role of the government is shifting from a service provider to a service enabler. For this reason, more attention should be paid to the risk management of indirect programs.

Second, welfare programs were outperformed overall by other programs, and achieved results that were indistinctive from other programs in terms of program necessity, adequacy of performance goals, monitoring and problem solving, enhancing efficiency, and conducting program assessment, although they obtained better results in budget execution and feedback.

In terms of redundancy, adequacy of the method to advance programs, and adequacy of performance goals, they achieved poorer results than other programs. The performance results indicate that redundancy occurred in welfare programs as increased welfare spending spurred competition between various line ministries and agencies to promote their own welfare programs, while prior planning was also in an inadequate

state.

Third, small-scale programs obtained poorer results than mid- or large-scale programs in setting performance goals, enhancing efficiency, achieving performance goals, and conducting program assessment. With regards to enhancing efficiency, greater budge size correlated with a greater opportunity for cost-cutting, and bigger programs were more likely to secure finances for assessing programs. As for proper performance planning, small-scale programs seemed to neglect performance management, suggesting that government agencies need to establish and operate a more effective management system.

Forth, ministry-managed programs were likely to perform poorly in most items. They received low ratings in performance planning, management and results. It should be noted that programs conducted by ministries were more politically motivated and faced more uncertain prospects in their future progress. In contrast, agency-managed programs had specific and clear objective or contents, often directly managed by the agencies themselves. According to regression analysis, however, even within the same category of direct programs, those operated by ministries yielded poorer results. Therefore, it is highly probable that organization characteristics correlated to performance results. Agency-managed programs were more likely to have a higher awareness and priority towards management performance, while ministries tended to neglect performance management as a formality. In particular, comparison of senior decision-makers' accountability and interest in performance management showed a greater gap between ministries and agencies.

The study results suggest that ministries should establish a proper management system, institute greater accountability for high-ranking officials, and utilize a performance management system. Currently, the executive branch in Korea operates various performance management and assessment systems, which are rarely used by senior officials for decision-making. Such systems are being perfunctorily managed because a decentralized responsible organization makes it hard to operate consistent systems established at a government level, and responsible organization or ministries lack the capability for dealing with performance management and assessment.

<Table V-1> Summary of SABP Analysis

	Indirect Program	Welfare Program	Economic Program	Small-scale Program	Large-scale Program	Ministry -managed Program
Total Score	- ***	- ***		- ***	+ **	- ***
Design	- ***	- ***	- ***	- *		
Management	- ***		+ ***			- ***
Results	- ***	- ***		- ***	+ ***	- ***
Clear Purpose	- ***		- **			
Spending Necessity						
Redundancy	- ***	- ***	+ *			+ **
Program Promotion Methods	- ***	- ***	+ **			
Performance Indicators	- ***			- *		- **
Goals	- ***	- ***	- ***	- *		
Monitoring	- ***				+ **	- ***
Problem Solving	- ***		+ ***			
Budget Execution		+ ***	+ ***	+ ***		- *
Efficiency Enhancement	- ***		- ***	- ***	+ *	- ***
Performance Goal Achievement 1	- ***	- ***		- *		- ***
Performance Goal Achievement 2	- *		+ *			- ***
Program Evaluation	+ ***		- **	- ***	+ ***	- **
Feedback	+ ***	+ ***	+ ***		+ **	- ***

Note: Relationships between two variables at each significance level are marked as follows: at levels less than 0.01 (***), 0.05 (**), and 0.1 (*), and coefficient direction is marked as +/-.

Considering the current and future trends of rising welfare spending, and the inadequacy of evaluation results in this sector, welfare programs were selected separately for additional analysis. The results of this analysis were obtained after categorizing welfare programs according to program purpose and matching additional SABP results. The analysis results below are obtained by controlling other possibly influencing elements.

- Mandatory spending programs obtained poorer performance

results than discretionary spending programs.

- Program performance showed a positive correlation to the budge size.
- Agency-managed programs outperformed ministry-managed programs.
- Voucher-supported programs were observed to receive favorable ratings in program design, management, and performance, though results were not significantly different according to support type (vouchers / goods / cash).
- In terms of program field, child-care and job creation programs obtained below-average performance results. Child-care programs showed below-average ratings in performance planning and achievement, job creation programs in budget execution.

It is interesting that mandatory spending programs in the welfare sector received low ratings in many items. They obtained poor performance in setting performance planning, budget execution, and program assessment seemingly because responsible line ministries / agencies neglected the programs due to the mandatory expenditure, which is not conducive to budge cuts. The low ratings in budgeting indicate that estimation for program demand would be incorrect or services were not properly delivered to beneficiaries. The results highlight the need for strengthening performance management of mandatory spending programs, making correct demand estimation, and improving a service delivery system.

The low ratings in performance management of mandatory spending programs suggest that systematically stable programs would not carry out effective management and improvement without regular inspection and pressure for system improvement. Again, it is highly likely that responsible line ministries / agencies take advantage of fact that mandatory spending programs are rarely subjected to budget adjustment or systemic improvement in reality even after being evaluated. Therefore, instead of handling mandatory spending programs with the existing evaluation method, the authorities need to seek a fundamental improvement including legislative amendment by utilizing an additional evaluation system on a

regular basis.

2 Conclusion and Directions for Future Research

This paper attempted to analyze performance results by program type and derive policy implications based on SABP results. Despite restrictions posed by limited date, the SABP derived the following results of significance:

- Programs that were directly managed by the government or services that were offered by the government outperformed indirect programs that the government consigned to other agencies.
- Welfare programs produced poorer results than other programs.
- Program performance showed a positive correlation to the budge size.
- Programs that were promoted by an agency-level organization earned favorable ratings.
- Mandatory spending programs obtained poorer evaluation results than discretionary spending programs.

In particular, the rapid expansion of welfare programs confirmed systemic improvements for the maintenance of the implementation structure and enhancement of performance results. This suggests that rather than focusing expanded financial resources, it is more important to establish and operate a performance management system corresponding to the increased budget.

The research results also show that programs consigned to the private sector or local governments should raise their efficiency by setting up a proper performance management system. With government programs increasingly outsourced, risks accompanied with outsourcing should be properly managed.

Small-budget programs are likely to face management issues due to budge limitation; however, the issue is not prominent enough to warrant consideration as a policy agenda. However, programs with small budgets

should strive to improve their performance by benchmarking other programs with similar characteristics, since ministries operate several similar small-scale programs.

Favorable ratings in agency-managed programs suggest that performance management capabilities can be different depending on characteristics of an operating organization. Although ministries may expend greater focus on developing political agendas than program management, growing political maturity in Korea should place the management and improvement of key programs at the center of the organizational culture.

The low ratings in performance management of mandatory spending programs suggest that systematically stable programs would not carry out effective management and improvement without regular inspection and pressure for system improvement. Again, it is highly likely that responsible line ministries / agencies take advantage of fact that mandatory spending programs are rarely subjected to budget adjustment or systemic improvement in reality even after being evaluated. Therefore, instead of handling mandatory spending programs with the existing evaluation method, the authorities need to seek a fundamental improvement including legislative amendment by utilizing an additional evaluation system on a regular basis.

Future research will develop the aforementioned policy implications into specific policy tasks. For instance, the expansion of welfare programs necessitates the identification of specific problems and the improvement thereof, by benchmarking favorable cases. In addition, co-management among welfare-related line ministries / agencies and a management system for service providers and beneficiaries should be established. Another important policy task is to enhance overall capacity for program management and to prepare an incentive system. In particular, as mentioned earlier, mandatory spending programs, small-scale programs, and ministry-managed programs failed to raise their performance because of a weak performance management system. This problem must be discussed in connection with the betterment of the governmental management system (Park and Brumpy, 2012).

<References>

Kim, John and Nowook Park, *Performance Budgeting in Korea*, OECD Journal on Budgeting, No.4, 2007.

Ministry of Strategy and Finance, *The 2011 Self-Assessment of Budgetary Program Manual*, 2012.

National Assembly Budget Office, *Overall Analysis of FY 2011 Settlement of Accounts*, 2012.

Park, Nowook, "Does More Information Improve Budget Allocation? Evidence from Performance-oriented Budgeting in Korea," Working Paper Presented at the Annual Congress of the International Institute of Public Finance, Dresden, Germany, 2012.

Park, Nowook and Jae Young Choi, *Making Performance Budgeting Reform Work*, Working Paper, World Bank Institute, forthcoming.

Park, Nowook and Jim Brumby, "Performance-oriented Budgeting in Korea: Overview, Challenges and Policy Agenda," Working Paper Presented at High-Level Conference at the Public Expenditure Management Network in Asia, Seoul, Korea, 2012.

Nowook Park is the founding director of the Center for Performance Evaluation and Management at the Korea Institute of Public Finance. He has served in many evaluation-related committees in various ministries and public organizations. He has been working as a consultant to international organizations including the World Bank and the World Intellectual Property Office. He has been an active researcher in the field of public financial management and program evaluation, and provided many presentations and training programs to government officials in various countries. He received BA in Economics from the Seoul National University and Ph.D. in Economics from the University of Michigan.

Jonghak Weon is a research fellow at the Korea Institute of Public Finance, Seoul, South Korea. His research interests include performance management and evaluation, delivery system of social welfare, labor market. Some of his recent works are *Policy Agenda for Budgetary Programs in the Koran Government* (KIPF 2012), *Graduating During the Recession and its Medium-term Implications on Public Finance* (KIPF 2012), and *A study on the Efficiency of Welfare Program (I): Focused on Local Government and NPO* (KIPF 2011). Weon has a Ph.D. in Economics from Hitotsubashi University.