

재정사업 자율평가의 정부 부처별 적합성 분석: 성과측정 적합도 개발과 적용을 중심으로*

김 영 록**

〈目 次〉

- I. 서론
- II. 사업·성과지표 분석 및 성과측정 적합도 개발
- III. 재정사업 자율평가의 단위 사업, 성과지표 분석
- IV. 결론

〈요 약〉

성과주의 예산제도(우리나라의 '재정사업 자율평가제도')의 핵심은 정책의 성과를 평가하여 그 결과를 예산 배분에 반영하는 것이다. 그렇다면 그 정책은 과연 측정 가능한 것인지, 또한 정책 평가를 위한 수단이라고 할 수 있는 성과지표는 측정 가능한 것으로 구성되어 있는지를 밝히는 것은 필수적인 과정이라고 할 수 있다.

따라서 본 연구는 재정사업 자율평가를 수행하는 전 부처의 정책 유형을 분류하고, 성과지표의 적절성을 파악하여, 이를 토대로 '성과측정 적합도'를 개발하였다. 또한 기관별 '성과측정 적합도'를 산출하여 성과측정이 용이한 기관과 그렇지 않은 기관을 수치화하였다.

본 연구를 통하여 이론적 논의를 기반으로 한 선행연구에서 시도했던 정책의 분류와 성과관리와의 관계를 실증적으로 규명하였다. 또한 시차를 두고 성과관리 적합도를 측정하여 기관들의 사업 수행 성격이 어떻게 변하여 가는지, 혹은 성과지표의 적합도가 향상되는 지를 관찰하는 것은 향후에 성과관리와 관련하여 시사점을 줄 것으로 예상된다.

【주제어: 성과주의 예산제도, 성과측정 적합도(PMA), 재정사업 자율평가, 정책 유형, 성과지표】

* 이 논문을 준비하는 과정에서 도움을 주신 연세대학교 행정학과 하연섭 교수님, 국무총리실 양홍석 과장님, 기획재정부 이강호 과장님, 최지영 사무관님, 국토해양부 김경욱 과장님, 서진희 서기관님, 교육과학기술부 안용환 사무관님께 깊은 감사의 말씀을 전합니다.

** 교육과학기술부 산하 한국교육학술정보원 선임연구원(rockim@keris.or.kr)
논문접수일(2010.7.26), 게재확정일(2010.8.27)

I. 서론

예산의 효율성을 향상하고 예산편성의 합리성을 제고하기 위한 예산의 성과 관리제도가 우리나라에 도입된 지 10여년이 지났다. 2000년부터 2004년까지의 시범사업기간을 제외한다면 본격적인 제도의 운영은 5년 이상의 시간이 경과하였다. 재정성과관리제도 또는 성과주의 예산제도(이하 성과주의 예산제도)가 도입 초기를 지났다고 볼 수 있다.

성과주의 예산제도의 핵심은 정책의 성과를 평가하여 그 결과를 예산 배분에 반영하는 것이다. 즉 정부 정책의 성과를 평가하여 그 결과가 우수한 정책에는 예산을 증액하고, 그렇지 않은 정책이나 사업은 예산을 삭감하는 것이다. 그러므로 평가 대상 정책이 과연 측정 가능한 것인지, 또한 정책 평가를 위한 수단이라고 할 수 있는 성과지표가 측정 가능한 지표로 구성되어 있는지를 밝히는 것은 성과주의 예산제도의 공과를 평가하는데 필수적인 과정인 것이다. 왜냐하면 성과주의 예산제도의 핵심은 성과 측정의 결과에 근거해서 예산 배분의 효율성을 높이는 것인데, 그 전제는 성과측정이 가능하다는 것이고, 측정가능한 정도 역시 동일하다는 것이다. 그러나 실제로 성과주의 예산제도와 다양한 정책의 영역들은 여러 가지 변수로 인해 적합한 정도가 상이하다(<표 1> 참조).

<표 1> 성과주의 예산제도에 영향을 주는 요인

내적 요인	외적 요인
-성과지표 등 성과정보의 적절성	-행정부에 대한 정치적 영향력
-프로그램의 특성과 평가와의 관계	-해당 국가의 통치운영 전반 특성
-피평가자 소양 및 이해도	-발생주의 회계 등 다른 제도와 연관성
-평가제도 자체의 완성도	-이익집단, 미디어 등 환경적 요인

출처 : 성과주의 예산제도와 관련된 선행연구를 바탕으로 연구자가 도출

이렇듯 다양한 변수로 둘러싸인 성과주의 예산제도를 효과적으로 분석하기 위해서는 논의 초점을 줄일 필요가 있다. <표 1>의 외적 요인에 대한 분석 방법은 별도로 하더라도, 성과주의 예산제도의 내적 요인들은 분석의 일반화가 어느 정도 가능할 것이다. <표 1>의 내적 요인을 중심으로 성과주의 예산제도를 분석할 수 있다면, 이 제도가 현실에서 실현되는 메커니즘을 보다 객관적으로 설명하는데 중요한 도구가 될 것이다. 성과주의 예산제도의 외적 변수를 제거한다면, 결국 성과주의 예산제도의 핵심은 측정 가능한 사업인가의 문제와 그 사업의 하위 성과지표가 측정가능한가라는 두 가지 핵심 변수로 집약된다. 왜냐하면 내적 요인 중 정책(프로그램)의 특성, 성과지표의 적절성은 성과주의 예산제

도 본연의 목표인 측정가능성(measurability) 여부에 대한 답을 줄 수 있는 가장 주요한 변수이다. 본 연구는 위 두 가지 변수(단위 사업, 성과지표)를 우리나라 재정사업 자율평가 보고서에 근거하여 분석한다. 이를 통해 각 기관별로 성과주의 예산제도와와의 적합 정도를 실증적으로 도출하여 정책 영역별로 차별적인 성과주의 예산제도의 설계가 필요함을 증명하도록 한다. 또한 이를 통해 성과주의 예산제도 도입과 설계, 그리고 제도 운용에 있어서 어떠한 고려가 필요한지를 밝히도록 한다.

이 연구는 다음과 같은 순서로 구성되어 있다. 먼저 제2장에서 ‘정책 영역과 성과주의 예산제도와와의 적합 정도’를 실증적으로 증명하기 위해 사업의 특성 분석과 성과지표 분석 기준을 마련하여 ‘성과측정 적합도’(performance measurement appropriateness, 이하 PMA)를 개발하였다. 이어 제3장에서는 우리나라의 재정사업 자율평가 결과(2008년도 기준, 2009년에 발표)에 근거하여 각 기관별 성과측정 적합도(PMA)를 도출하고 그 의미를 논의하고 있다. 끝으로 제4장에서는 논의를 요약하고 정책적 시사점을 도출하도록 한다.

II. 사업·성과지표 분석 및 성과측정 적합도 개발

1. 성과측정 적합도의 정의

제1장에서 논의한 바와 같이 성과주의 예산제도의 외적 변수를 제거한다면, 결국 성과주의 예산제도가 제대로 기능하기 위해서는 측정 가능한 프로그램인가의 문제와 그 프로그램의 하위 성과지표가 측정 가능한가라는 두 가지 핵심 변수로 집약된다.

<표 2> 정책(프로그램)의 특성과 성과지표의 적절성 분류기준

프로그램의 유형성 여부	성과지표의 적절성 여부
-프로그램의 유형성 : SOC, 자격증 발급 등과 같은 프로그램의 유형성 여부	-객관적 지표 여부 : 수치로 성과목표 제시 또는 물리적 측정가능한 기술 여부
-프로그램의 무형성 : 교육, 보건 등과 같은 무형적, 이념적 프로그램	-결과측정여부 : 성과지표 설정이 단순한 작업량(workloads-oriented) 측정이 아닌 결과 측정과의 연계여부

출처: Pollitt(1999), Bouckaert & Ulens(1998), Chun & Rainey(2005)에서 발췌 및 재구성

따라서 본 연구에서는 성과주의 예산제도에 영향을 줄 수 있는 요인들 중 정책의 특성과 성과지표의 특성을 핵심 변수로 설정한다. 이 두 가지 핵심 변수를

기준으로 분석을 실시하여 성과측정 적합도(PMA)를 개발하도록 한다. ‘성과측정 적합도’란 각 기관이 수행하고 있는 사업의 성격을 최종 결과물(outcome)의 성격을 중심으로 유형성 여부(tangibility)를 판단하고, 사업 성과지표의 객관성 정도, 결과 측정여부를 평정하여 현재 시점에서 성과관리에 얼마나 적합한지를 보여주는 지표라고 정의할 수 있다. 사업의 성격을 평정하는 것은 각 기관이 수행하고 있는 사업의 고유한 성격이 성과관리에 얼마나 적합한지를 판단하기 위함이고, 성과지표를 분석하는 이유는 각 기관의 성과지표가 성과관리에 적합하도록 개발되었는지를 판단하기 위함이다.

이를 위하여 각 기관의 사업 또는 정책의 유형성을 평정하는 작업과 성과지표를 분석하는 작업을 실시할 필요가 있는데, 먼저 각 단계에 해당하는 정의와 하위 지표를 명확히 할 필요가 있다. 다음 절에서는 선행연구에서 포괄적으로 제시된 기준과 논의를 중심으로 세부적인 기준을 설정하도록 한다.

2. 사업의 특성 분석

사업(프로그램)¹⁾의 특성과 관련하여 분류하는 논의는 Pollitt(1999)과 Bouckaert 외(1998) 등의 프로그램 구분이 매우 큰 시사점을 준다고 할 수 있다. Pollitt은 재무관리와 성과관리의 통합 가능성을 다음과 같은 기준으로 분류하였다. 그 기준은 유형적(tangible)으로 측정가능한가, 혹은 표준화되고 구체적인 서비스로 구성된 프로그램인가, 또는 프로그램의 효과 측정에 소요되는 시간 등이다(당연히 프로그램의 효과 측정에 장시간이 소요될 경우 재무관리와 성과관리는 힘들다). Bouckaert 외(1998) 역시 프로그램의 유형성(tangibility) 여부를 기준으로 프로그램을 3가지로 구분하였다. Bouckaert 외(1998)는 프로그램의 유무형을 기준으로 세 가지 타입(type)으로 분류될 수 있다고 보고, 건설과 같은 영역을 ‘Type 1’로, 교육과 같은 영역을 ‘Type 2’로, 보다 무형적이고 이념적인 영역을 ‘Type 3’으로 규정한 바 있다. 유형적(tangible) 서비스의 예로 쓰레기 수거, SOC 건설, 공공도서관 운영 등을 예로 든 바 있다. 이러한 서비스들은 무형적 서비스에 비해 보다 성과 측정이 용이한데, 그 이유는 산출물을 보다 쉽게 측정할 수 있기 때문이다. Type 2에 속하는 서비스의 영역은 주로 인간과 관련된 서비스로 교육, 장애인정책, 보육, 노인 프로그램 등이다. 이러한 프로그램은 Type 1에 비

1) 본 연구에서는 프로그램, 사업, 또는 단위 사업, 정책 영역 등의 용어들이 맥락적 동질성을 가지고 사용되지만, 우리나라 재정사업 자율평가제도, 프로그램 예산제도 등 성과관리제도, 예산제도에서 사용되는 용어들의 제도적 의미는 차이가 있다. 즉 프로그램은 동일한 정책 목표를 달성하기 위한 단위사업(activity/project)의 묶음이며, 정책적으로 독립성을 지닌 최소 단위를 말한다. 또한 우리나라가 채택하고 있는 프로그램 예산제도의 기본 구조는 일반적으로 「정부기능(function) - 정책(policy) - 프로그램(program) - 단위사업(activity)」의 계층구조를 갖는다(박노옥, 2008: 35-36).

해 평가지표가 훨씬 섬세하여야 함은 물론, 그 효과 측정에 많은 시간이 소요된다. Type 3은 프로그램이 아이디어(idea)적 성격이 강한 것을 지적하였다. 정부의 대국민 캠페인 등이 이에 해당한다. Bouckaert 외(1998)의 분류는 정부 프로그램의 분류를 유무형성을 기준으로 시도하였다는 점에서 의의가 있다. 그러나, 그 기준이 직관적이고, 구체적이지 않다는 한계가 있다.

또한 실질적 정책(substantive policy)과 절차적 정책(procedural policy)으로의 정책 분류(Anderson, 1975; Sabatier, 1991: 147-148)도 정책 산출물의 특성이 주요 기준이 되는 분류이다. 아울러 Vedung 외(2003: 28)의 분류 역시 본 연구의 주요한 논거를 제공한다. Vedung 외는 유형적 자원(tangible resources) 또는 실질적 권력을 사용하여 이루어지는 정책을 물질적 정책(Material Policies), 무형적인 이익 또는 불이익을 배분하는 정책을 상징적 정책(Symbolic Policies)으로 구분하였다.

한편, 정책 수단(policy tool) 관점에서 분류를 시도했던 Salamon(2002)은 유사한 개념으로 정책 수단의 ‘가시성(visibility)’이라는 개념을 도입하였다. 정책의 가시성이란 일반적인 정부 재정과 정책 검토 과정에서 도구에 투입된 자원들이 보여주는 정도를 말한다. 가령 유아 교육 분야에서 보다 가시적인 수단(tools)은 보조금(grant), 바우처(voucher) 등이 될 수 있으며, 가시적이지 않은 수단은 규제, 세금 감면 등이라고 지적하였다. Salamon의 구분은 정책 수단의 성격에 초점을 둔 것으로, 예산의 가시적인 정도에 따라 정책 수단들을 배열하였다.²⁾ 그러나 Salamon의 구분은 프로그램 자체의 가시성이나 물리적인 특성 여부보다는 예산이 직접적으로 투입이 되었는가 여부에 집중하기 때문에 Pollitt이나 Bouckaert의 논의와는 차이가 존재한다.

또한 조수현(2007: 64-65)은 정책의 영역 분류를 Pollitt(1999: 18-19)의 논의를 기초로 발전시켰는데, 사업의 재량성과 성과관리의 수월성이라는 두 가지 기준으로 정책영역 구분을 시도하였다. 특히 정책의 유형적인 산출물을 기준으로 정책 영역을 분류하였다. 그런데, Pollitt의 분류와는 달리 물리적인 산출물이 없어도, 보건분야 예산을 예로 들어 사업결과 산출이 명료하다면 성과측정이 가능한 것으로 보았다.

요약하자면, 정책 또는 프로그램의 특성에 따른 분류와 관련한 선행 연구를 바탕으로 이 연구에서 성과주의 예산제도에 영향을 줄 수 있는 핵심 요인으로서 정책의 특성, 즉 유형성 여부를 활용하기 위해 세부 정의를 할 필요가 있다. 유형성에 대해서 선행 연구자들의 기준처럼 그 범위를 좁게 해석할 수도 있고 넓게 해석할 수도 있지만, 본 연구에서는 최종 결과물에 대한 논란의 여지가 있는 사업들을 배제하기 위해 최대한 보수적으로 해석하고자 한다.

2) Salamon은 visibility 외에도 directness, coerciveness, automaticity 등의 기준을 제시하였다 (Salamon, 2002).

프로그램이 유형적인지를 판단하기 위해 가장 먼저 고려해야 할 사항은 정책(프로그램)의 최종 생산물이나 활동의 결과물이 물리적으로 확인 가능 하느냐 여부이다. 정부 정책 중 교량 건설, 도로 유지 보수 등 SOC 관련 프로그램이 대표적인 예이다. 또한 각종 장비 현대화, 자격증 인증 사업과 같이 객관적으로 데이터 산출이 가능한 프로그램 역시 유형적인 영역에 속한다고 본다. 예를 들어 지식경제부가 추진하고 있는 프로그램인 ‘국가표준혁신’ 사업(관련 성과지표 : 국가표준(KS) 제정 건수 등)은 논란의 여지가 없는 객관적 데이터로 판단 가능한 영역이다.

〈표 3〉 정책(프로그램)의 유형성에 대한 정의 및 분류 기준

종류	정의 및 예시
유형적 (tangible)	-정부 프로그램의 최종 생산물(product) 또는 활동(activity)의 결과물이 물리적으로 인지 가능함 예)신항만 건설 사업, 인공위성 개발 사업 -측정가능하고 표준화되어 있는 활동으로서, 반복적인 재화나 서비스를 제공하는 사업 예) 국가표준혁신 사업(관련 성과지표 : 국가표준 KS 제정 건수 등)
무형적 (intangible)	-정부 프로그램의 최종 생산물(product) 또는 활동(activity)의 결과물이 물리적으로 인지 불가능함 예)대학자율역량 기반조성 사업, 인문사회 학술연구조성 -표준화하기 어렵고, 일상화되기 어려운 서비스를 제공하는 사업 예)교육과정 개정 및 후속지원 사업(성과지표 : 교과과정 해설서 이용자 만족도), 기초연구개발사업, 정책아이디어제공 등

출처: Pollitt(1999a: 18-19), Bouckaert 외(1998) 및 조수현(2007: 64)의 논의를 참고하여 작성함

a. Pollitt(1999a)과 Bouckaert 외(1998)는 유형적인 기준을 물리적 인지에 초점을 둔 반면, 조수현(2007)은 사업결과의 측정가능성에 기반하여 성과관리의 용이함을 논하여 유형적인 정책영역의 범위를 확장함. 그러나 본 연구는 유형성의 범위를 최대한 보수적으로 해석하여, 물리적 인지 여부에 초점을 맞춤. 그 이유는 정책의 유무형성을 판단하는 작업 이후 성과지표를 분석하는 2차 작업이 있기 때문에 논란의 여지가 될 수 있는 영역의 정책은 성과지표 분석을 통해 더욱 명확히 판단할 수 있기 때문임.

반면 정부 프로그램 중 최종 생산물이나 활동의 결과물이 물리적으로 확인할 수 없는 영역은 무형적이라고 규정한다. 가령 교육과학기술부의 2008년도 프로그램 중 하나인 ‘평생학습 기반구축 및 평생학습 활성화지원’은 물리적으로 확인 가능한 최종 결과물을 찾을 수 없다(이 경우 성과지표는 수치로 확인할 수도 있지만, 그것은 다음 절의 ‘성과지표의 적절성’에서 판단하도록 한다). 또한 데이터로 표현할 수는 있지만, 평가자에 따라서 결과가 달라질 여지가 있는 프로그램도 무형적이라고 규정한다. 예를 들어 국방부의 ‘예비군 교육훈련

성과 제고'는 최종 생산물이 물리적으로 확인이 불가할 뿐만 아니라, 성과가 수치로 표현된다고 하여도 평가자에 의해 달라질 수 있기 때문에 무형적 영역에 속한다고 본다. 성과지표를 수치화하기 위해 만들어진 만족도 조사, 개선지수, 적정률 등은 주관적일 수 있는 여지가 있다. <표 3>은 정책(프로그램)의 유형성 여부에 대한 정의와 기준, 예시이다.

3. 성과지표의 분석

성과지표 역시 앞서 살펴본 프로그램의 성격과 함께 성과주의 예산제도에서 핵심적인 사항이다. 그동안 성과지표의 적절성과 관련된 논의는 매우 다양하게 전개되어 왔다. 최근 연구 중에서는 Robinson(2007: 26-27)의 구분은 성과주의 예산제도를 구성하는 기본적 핵심 개념을 명확하게 정의하였는데, 그 가운데 결과물(outcomes)과 산출물(outputs)의 구분은 본 장의 성과지표 분석에 유용하다. Robinson에 의하면 결과물은 산출물들의 의도된 효과이다. 예를 들어 교통사고 시 환자에 대한 처치는 산출물이며, 의도된 결과물은 환자의 생명 유지 혹은 장애의 최소화인 것이다. 또한 Robinson은 결과물을 '근접 결과물'(proximate outcomes)과 '고급 결과물'(high-level outcomes)로 구분하기도 하였다. 고급 결과물이 산출물의 공급 목적 또는 궁극적 목적을 의미한다면, 근접 결과물은 산출물의 보다 직접적 혹은 즉각적 영향이라고 볼 수 있다. 학교 교육에 있어서 지식이 '근접 결과물'이라면, 이로 인해 기대할 수 있는 '경제 발전'은 고급 결과물인 것이다.

또한 성과지표의 적절성을 분석하기 위해 '목표 모호성(goal ambiguity)' 관련 연구 중 평가적 목표 모호성(evaluative goal ambiguity)을 밝히기 위해 실증적으로 성과지표를 분석한 툄(Chun & Rainey, 2005; Lee, Rainey & Chun, 2009) 역시 유용하다. 평가적 목표 모호성(evaluative goal ambiguity)은 어떤 조직의 사명을 얼마나 달성하였는지 그 진전을 평가하는데 있어서 발생하는 경쟁적 해석의 정도를 말한다. 조직의 성과를 평가할 경우 조직의 사명은 성과지표(performance indicators) 혹은 성과표적(performance targets)으로 전환되어야 한다(Grizzle, 1982; 전영환, 2004: 56).

성과지표를 어떻게 설정했느냐는 해당 기관의 성과주의 예산제도 수행 가능성을 좌우한다. 왜냐하면 성과지표가 객관적(objective)이고 결과지향적(result-oriented)인가 혹은 주관적(subjective)이거나 작업량 지향적(work-oriented)인가에 따라 성과가 측정가능한지 또는 그것이 유의미한지의 척도가 될 것이기 때문이다. 따라서 성과지표를 분석하는 것은 성과주의 예산제도를 분석하는데 필수적인 요소라고 볼 수 있다. 그리고, 다양한 기법이 있을 수 있겠지만 본 연구에서는 평가 모호성 연구에서 사용되었던 성과지표 분석 기준을 사용하도록 한다

(Chun & Rainey, 2005; Lee, Rainey & Chun, 2009). 그 이유는 평가적 모호성을 분석하기 위한 성과지표 분석 기준이 개념적으로 성과관리의 적합성을 분석하기 위한 본 연구의 목적과 매우 일치할 뿐만 아니라, 미국 연방정부를 구성하는 각 기관의 성과계획서(performance plans)와 성과보고서(performance reports)를 실증적으로 분석하여 타당성을 확보하였기 때문이다. 또한 실제 성과지표 평정 작업시 Robinson의 기준을 보충적으로 사용하도록 한다. 평가적 모호성 개념을 측정하기 위해 고안된 성과지표의 분류에서 작업량 지향적 성과지표(workload oriented indicators)는 Robinson의 구분을 적용하면 산출물(output)과, 결과지향적 성과지표(results-oriented indicators)는 근접 결과물(proximate outcomes) 및 고급 결과물(high-level outcomes)과 맥락적 동질성을 갖는다고 볼 수 있다.

<표 4>는 성과지표를 분석하기 위한 정의 및 기준을 기술한 것이다.

<표 4> 성과지표의 분류 정의 및 기준

종류	정의 및 예시
-객관적 지표 (objective indicators)	-계량적 성과목표 제시 또는 물리적으로 관찰 가능한 상황 또는 측정가능 기술 여부 예) 교통 사고 사망자 수
-주관적 지표 (subjective indicators)	-계량적 표적수준을 제시하지 않으며, 개인의 인지에만 의존하는 성과지표 예) 교과과정 개선안 마련 완료
-작업량 지향 지표 (Workload-oriented indicators)	-결과(outcome) 지표 혹은 효율성(efficiency) 지표를 제외한 입력(input) 및 산출(output)지표. 즉 결과 및 생산성 측정과 연관이 없는 단순 작업량 예) 교육정보자료공유 현황(프로그램명 : 교수학습 혁신)
-결과지향 지표 (Results-oriented indicators)	-결과(outcome) 및 효율성(efficiency)를 측정한 지표. 즉 결과 및 생산성 측정과 연계되는 지표 예) 교도소내 훈련참여자의 3년내 재복역율(프로그램 : 교도작업운영 및 직업훈련)

<표 4>는 다시 아래의 표를 통해 성과지표를 분류할 수 있을 것이다.

<표 5> 성과지표의 분류 형식

	subjective(S)	Objective(O)
Workload-oriented(W)	WS	WO
Results-oriented(R)	RS	RO

출처 : Lee, Rainey and Chun(2009)에서 정리

- a. WS : 작업량중심적이면서 주관적 지표 / WO : 작업량중심적이면서 객관적 지표
- RS : 결과지향적이면서 주관적 지표 / RO : 결과지향적이면서 객관적 지표

<표 4>의 구분과 같이 성과지표는 객관적 지표, 주관적 지표, 결과지향적 지표, 작업량 지향적 지표로 구분된다. 주관적 성과지표는 계량적 표적 수준을 제시하지 않으면서 개인의 지각에 의존하는 지표를 말한다. 반대로 객관적 지표는 계량적 목표 수준을 제시하는 지표이지만, 지표에 따라서 반드시 계량적 수준을 제시할 필요는 없다. 성과지표의 객관성은 달착륙과 같은 실체적 조건이나 사건을 관찰함에 의해서도 확보될 수 있기 때문이다. 미국 정부성과법(GPRA)에 따르면 성과지표들은 계량적이거나 혹은 최소한 객관적이어야 한다고 규정하고 있지만 실현 불가능할 경우 OMB(관리에산처)의 승인을 얻어 주관적인 성과지표를 쓸 수 있도록 예외를 두고 있다(전영한, 2004: 60).

한편 <표 4>와 <표 5>에 의하면 하나의 성과지표는 두 번의 평정 작업을 거치게 된다. 먼저 <표 4>의 예시에 있는 ‘교통사고 사망자 수’ 라는 성과지표는 객관적 지표인지 혹은 주관적 지표인지 여부를 평정하게 된다. 정의에 의해 계량적 성과목표를 제시하거나 물리적으로 명확히 관찰 가능한 지표이기 때문에 객관적 지표로 평정받게 된다. 아울러 이 성과지표가 성과목표를 달성하기 위한 결과지향적 지표(results-oriented indicators)인지 혹은 단순히 작업량을 측정하기 위한 지표(workload-oriented indicators)인지를 평정하게 된다. 가령 ‘교통사고 줄이기’ 라는 성과목표가 있다면 ‘교통사고 사망자 수’는 결과지향적 지표로 평가받을 수 있다. 성과지표가 성과목표가 직접적으로 연동되기 때문이다. 이렇게 하나의 성과지표가 두 차례 평정 과정을 거쳐서 <표 5>의 기준에 의해 구분될 수 있다. 가령 ‘교통사고 사망자 수’ 라는 성과지표가 객관적(Objective)이면서, 결과지향적(Results-oriented indicators) 지표로 평정된다면 <표 5>의 ‘RO’로 표기될 수 있을 것이다. 이러한 방식으로 각 기관의 성과지표를 분석한다면 <표 5>의 네 가지 구분으로 성과지표는 나누어질 수 있을 것이다. 성과관리의 특성을 고려하였을 때 성과지표가 RO(결과지향적 지표이면서 객관적 지표)의 성격을 띠면 성과관리에 당연히 용이할 것으로 유추할 수 있다. 반대로 성과지표가 WS(작업량 지향적 지표이면서 주관적 지표)로 평정된다면 성과관리에 어려움이 따를 것이다.

4. 성과주의 예산제도와 프로그램, 성과지표 분석을 위한 모형

<표 3>와 <표 4>에 기술된 내용을 기준으로 정책 영역은 그것들을 구성하는 프로그램의 성격(유무형성)을 분석하고, 성과지표들의 적절성을 파악함으로써 성과주의 예산제도와 어떻게 부합하는지 아래 <그림 1>와 같이 기술될 수 있을 것이다.

<그림 1> 성과주의 예산제도와 정책 프로그램별 적합도

유형적 프로그램	
<1사분면>	<2사분면> 적합도 上 예)SOC, 자격인증, 체포율
주관적/작업량 지향 성과지표(WS)	객관적/결과지향 성과지표(RO)
<3사분면> -적합도 下 예)교육 정책, 이념적 서비스 등	<4 사분면>
비유형적 프로그램	

<그림 1>과 같이 각 정책 영역을 구성하고 있는 프로그램의 유형성 여부와 성과지표의 객관성과 결과측정 여부를 기준으로 2사분면에 위치한 정책 영역들이 가장 성과주의 예산제도와 부합도가 높다고 볼 수 있다. 반대로 3사분면에 위치한 정책영역들은 성과주의 예산제도가 적용되는데 별도의 고려사항이 필요한 영역이라고 볼 수 있다. 1, 4 사분면에 위치하는 정책영역은 성과지표와 프로그램의 성격이 혼재되어 있어서 명확한 해석이 힘든 영역이다.

즉, 성과주의 예산제도는 프로그램이 유형적(tangible)이고, 성과지표가 객관적(objective)이며 결과지향(result-oriented)으로 구성될 때 가장 효과적으로 작동한다. 이러한 가정은 다음의 공식으로 표현될 수 있다.

$$PB = f(TP, CPI, others)$$

PB = Performance Budgeting(성과주의 예산제도)

TP = Tangibility of Programs(프로그램의 유형성 여부)

CPI = Characteristics of Performance Indicators(성과지표의 성격 분류)

다음 절에서 우리나라에서 시행되고 있는 재정사업 자율평가 보고서 상의 프로그램과 성과지표를 위 기준에 의거해 분석하여 보다 실증적 결과를 확인할 수 있을 것이다.

Ⅲ. 재정사업 자율평가의 단위 사업, 성과지표 분석

앞 절에서 개발된 성과측정 적합도를 2009년 7월에 공표된 ‘2008년 재정사업 자율평가 결과’에 수록된 단위 사업, 성과목표, 성과지표에 근거하여 분석하도록 한다. 분석대상은 2008년도 기준 재정사업 자율평가를 실시한 33개 중앙행정기관의 단위 사업과 성과지표이다.

1. 분석 대상 및 방법

2008 회계연도 기준으로 재정사업 자율평가 대상기관은 모두 33개 중앙행정기관이며, 이들 기관이 제출한 일반재정사업 자율평가 최종보고서에 기재된 단위 사업과 성과지표들을 분석대상으로 삼았다³⁾.

분석 방법은 33개 기관이 제출한 최종보고서의 단위 사업과 성과지표를 모두 망라하여 2장에서 논의하였던 단위 사업의 유무형 여부와 성과지표의 분류를 시도한다. 이를 통해 2008년도 현재 우리나라 중앙행정기관 별로 성과주의 예산제도의 부합정도를 판별하여 <그림 1>의 어느 사분면에 위치하는 지 밝히도록 한다. 또한 단위 사업의 유무형성, 성과지표 분류의 타당성을 확보하기 위해 평정자간 신뢰도(inter rater reliability)를 측정하여 상관관계 계수를 확보하도록 한다. 평정자간 신뢰도를 측정하는 연구 방법은 Krippendorff(2004)에 의해 정의된 바 있다. Krippendorff에 의하면 평정자간 신뢰도를 확보하기 위해서는 ‘적어도 두 명 이상의 독립적인 평정자’를 필요로 한다고 하였다. 또한 신뢰도를 측정하는 지수(indices)는 여러 가지가 있으나 본 연구에서는 확률 일치도(percent agreement)⁴⁾를 사용하도록 한다⁵⁾. 또한 Chun & Rainey(2005) 역시 두 명의 독립적인 평정자의 평정 결과를 통해 신뢰도를 확보한 후 확률 일치도를 산출하기 위해 평정 결과 간의 상관관계 계수를 구하여, 상관관계를 확인하였다. 이후 상관관계가 명백하다면 두 개의 평정 결과 중 하나를 채택하여 분석하도록 한다.

3) 세부적인 기관별 단위사업과 성과지표 목록, 분석결과 등은 연구자에게서 입수 가능함

4) 확률 일치도(percent agreement)는 모든 평정자간의 일치도를 퍼센트로 나타내는 방법으로 상관 관계 계수를 구하여 일치도를 판단한다. 평정자간 신뢰도를 평가하는데 가장 일반적이면서 손쉬운 방법이지만 이 방법의 가장 중요한 전제는 평정자들이 평정 주제 또는 대상에 대하여 일정 수준 이상의 지식을 공유하거나 연수가 전제되어야 한다는 것이다(Lombard 외, 2002: 590-591). 따라서 본 연구에서는 평정자를 이러한 점을 고려하여 선정하였음을 밝힌다.

5) 평정자간 신뢰도(inter-rater or intercoder reliability)를 나타내는 방법에는 확률 일치도 (Percent Agreement), Holsti's Method, Scott's Pi(π), Cohen's Kappa(κ), Krippendorff's Alpha 등이 있다. 관련 상세 논의는 Lombard 외(2002) 참조.

본 연구에서는 연구자 외에 별도로 2차 연구자로 지정하여 독립적으로 단위 사업 및 성과지표를 평정하도록 하였다. 본 분석을 위하여 통계 프로그램은 STATA 10.0을 사용하였다.

2. 분석 대상 기술 통계 분석

재정사업 자율평가 결과보고서를 제출한 33개 중앙행정기관의 총 단위 사업 수는 모두 265개이며, 단위 사업을 평가하기 위해 설정된 성과지표는 모두 584개이다(기금 사업은 제외). 본 연구는 이들 265개 단위 사업과 584개의 성과지표를 제출한 <표 3>, <표 4>, <표 5>에 의거해 분류하였다. 분류한 결과는 <표 6>과 같다.

<표 6> 우리나라 중앙행정기관 단위 사업 및 성과지표 분류 및 기술통계치

부처명	단위사업수	성과지표수	O	R	tan	RO	WS
관세청	2	5	5	5	1	5	0
국세청	3	4	3	2	1	1	0
금융위원회	1	3	2	0	0	0	1
기획재정부	3	7	3	3	0	0	1
농림수산식품부	29	61	56	43	19	38	0
산림청	6	12	6	7	4	4	3
중소기업청	10	27	17	19	2	12	3
지식경제부	22	59	51	47	11	42	3
통계청	6	14	9	4	1	2	3
특허청	5	17	9	10	2	8	6
교육과학기술부	22	56	33	33	8	16	6
국가보훈처	3	7	3	2	0	1	3
국토해양부	18	39	37	32	16	32	2
기상청	3	5	4	5	3	4	0
노동부	5	13	6	8	0	5	4
농촌진흥청	5	12	6	4	0	3	5
문화체육관광부	21	51	41	22	9	21	9
문화재청	11	30	12	11	3	4	11
방송통신위원회	4	12	8	1	1	1	4
보건복지가족부	18	32	25	10	6	9	6
식품의약품안전청	3	5	4	4	1	4	1
여성부	3	7	4	0	1	0	3
환경부	15	23	19	10	7	10	4
경찰청	4	6	3	2	0	1	2
국방부	9	14	9	8	5	5	2
법무부	6	11	9	4	1	2	0
병무청	1	2	1	1	0	0	0
소방방재청	2	4	2	2	0	0	0
외교통상부	4	9	5	0	0	0	4

(다음 쪽에 계속)

부처명	단위사업수	성과지표수	O	R	tan	RO	WS
조달청	5	9	7	7	3	6	1
통일부	2	2	0	1	0	0	1
해양경찰청	3	5	5	5	3	5	0
행정안전부	8	21	9	10	2	1	3
합계	265	584	413	322	110	242	91
평균	8.03	17.70	12.52	9.76	3.33	7.33	2.76

- a. O: 객관적 지표(objective indicators), R: 결과지향적 지표(result-oriented indicators)
- b. 주관적 지표(subjective indicators) 및 작업량 지향적 지표(workload-oriented indicators)는 총 지표수에서 반대 영역에 해당하는 객관적 지표(O), 결과지향적 지표(R)의 숫자를 빼면 산출된다.
- c. tan: 유형적(tangible) 단위 사업의 수(비유형적 단위 사업의 수는 '단위 사업의 수에서 유형적 단위 사업의 수를 제한 값)
- d. OR: 객관적이고 결과지향적인 성격을 가지고 있는 지표의 수, WS : 주관적이고 작업량 지향적인 성격을 가지고 있는 지표의 수

<표 6>는 33개 중앙행정기관들의 단위 사업의 유무형성과 성과지표를 평정한 결과이다. 이를 위해 앞서 기술한 바와 같이 두 명의 평정자가 독립적으로 평정한 다음 상관관계(correlation)를 계산하여 신뢰도를 확보하였다.

<표 7>에서 보는 바와 같이 평정자간 평정결과의 상관관계는 평정 작업별로 상이하게 나타났다. 우선, 단위 사업들의 유무형성을 평정하는 작업 결과 평정자간의 상관관계는 매우 높은 것으로 나타났다(0.71, $p < .5$). 또한 성과지표의 객관적, 주관적 지표의 평정은 가장 높은 상관관계를 나타내었다(0.88, $p < .05$). 또한 성과지표의 작업량 지향적(workload-oriented indicators), 또는 결과지향적 지표(result-oriented indicators)를 평정하는 상관관계는 0.65로 통계적으로 유의미하게 나타났다⁶⁾. 이러한 결과가 의미하는 바는 두 명의 평정자가 실시한 세 가지 평정작업, 즉 사업의 유무형성 평정, 성과지표의 객관성/주관성 평정, 성과지표의 결과지향성/작업량 지향성 평정이 통계적으로 의미 있는 일관성을 보인다는 것이다. 이러한 결과를 바탕으로 본 연구는 두 가지 평정 결과 중 주 연구자의 평정값을 기준으로 기술하도록 한다.

6) 단순한 산출물(output)인지, 아니면 결과(result 또는 outcome)인지 여부를 평정하는 것은 결코 쉬운 일이 아니다. 성과목표와의 연계성을 면밀히 검토하여야 하고, 평정자간에 인식 차이가 존재하기 때문에, 객관적 지표 여부를 평정하였던 상관관계보다 결과지향적 지표 여부를 평정하였던 상관관계가 떨어진 것으로 판단된다.

〈표 7〉 변수별 및 평정자간 상관관계

	so_g	wr_g	tan_g	so_y	wr_y	tan_y
so_g	1.0000					
wr_g	0.2262*	1.0000				
tan_g	0.1591*	0.5042*	1.0000			
so_y	0.8834*	0.2381*	0.1769*	1.0000		
wr_y	0.0571	0.6594*	0.3917*	0.1081*	1.0000	
tan_y	0.2753*	0.4905*	0.7126*	0.2989*	0.3943*	1.0000

*p<.05

- a. so_y : 주평정자의 성과지표 객관성/주관성 평정값
- b. so_g : 부평정자의 성과지표 객관성/주관성 평정값
- c. wr_y : 주평정자의 성과지표 결과지향성/작업량 중심성 평정값
- d. wr_g : 부평정자의 성과지표 결과지향성/작업량 중심성 평정값
- e. tan_y : 주평정자의 사업 유무형성 평정값
- d. tan_g : 부평정자의 사업 유무형성 평정값

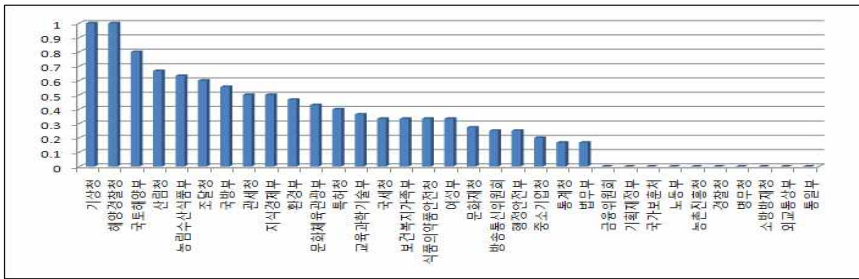
한편 <표 6>에 의하면 우리나라의 중앙행정기관들의 단위 사업의 수는 평균 8.03개이며, 성과지표의 수는 평균 17.70개이다. 따라서 하나의 단위 사업 당 성과지표의 수는 약 2.2개이다. 또한 성과관리 지표로서 제시한 단위 사업과 성과지표들을 1차적으로 평정한 결과 전체 성과지표들의 약 70.7%가 객관적인 지표로 구성되어 있으며, 약 29.3%가 주관적 지표로 구성되어 있었다. 따라서 우리나라 부처들이 설정한 성과지표들은 객관적 지표의 비중이 높은 편이라고 분석할 수 있다. 반면 성과지표 중 결과지향적(result-oriented) 지표의 비율은 55.1%, 작업량 지향(workload-oriented) 지표는 44.9%로 조사되었다. 따라서 성과지표들이 측정가능한 성과지표의 구성 비율(객관적 지표)은 높다고 볼 수 있으나, 단순 산출물(output)이나 작업량 측정을 위한 지표의 구성 비율(작업량 지향 지표) 역시 비교적 높다고 볼 수 있다. 한편 가장 성과관리에 적합한 성과지표라고 할 수 있는 객관적이면서, 결과지향적인 성과지표(RO)의 수는 242개로 전체 지표 중 41.4%를 차지하였다. 또한 성과관리에 가장 적합하지 않다고 할 수 있는 주관적이면서, 작업량 지향적인 성과지표(WS)의 수는 91개로 15.6%를 차지하였다. 43%에 해당하는 나머지 성과지표는 객관적 지표이기는 하나 작업량 지향적 지표(WO)이거나, 주관적 지표이기는 하나 결과지향적 지표(RS)이다. 이러한 성과지표들은 지표의 특성이 혼재되어 있는 지표로서 본 연구의 분석 대상에서 제외하기로 한다. 지표의 성격이 혼재되어 있다는 의미는 평정자에 따라서 결과가 달라질 수 있는 지표일 가능성이 있고, 지표의 특성이 분명한 지표를 대상으로 한 분석 작업이 각 기관의 성과지표와 성과관리와의 관계를 분명히 드러낼 수 있기 때문이다.

또한, 전체 265개 단위 사업 중 110개의 사업의 유형적(tangible)인 사업으로 평정되었으며 이는 전체의 41.5%에 해당한다.

3. 기관별 분석

앞 절에서 기술된 기술 통계치를 기반으로 이 절에서는 기관별 평정을 실시하였다. 본 절에서는 기관별로 <표 3>과 <표 4>에 근거하여 단위 사업과 성과지표를 평정하였다. 그 결과가 <그림 2>과 <그림 3>와 같은 결과가 도출되었다.

<그림 2> 단위 사업의 유형성(tangibility) 비중이 높은 기관 순서



<그림 2>는 기관별로 단위 사업의 유형성(tangibility)를 평정한 후 그 비율이 높은 순으로 나열한 것이다. 그 결과 기상청, 해양경찰청, 국토해양부, 산림청 등의 단위 사업이 유형성이 높은 것으로 나타났다. 즉 이들 기관의 수행 사업은 사업의 최종 생산물 또는 활동의 결과물이 물리적으로 인지가 가능하다는 의미이다. 반대로 금융위원회, 국가보훈처, 노동부, 외교통상부, 통일부 등은 실시하는 단위 사업의 유형성이 거의 관측되지 않았다. 다시 말하면 이들 기관이 수행하는 사업은 사업의 최종 생산물 또는 활동의 결과물이 물리적으로 인지가 불가능하다는 의미이다.

<그림 3> 객관적 지표 및 결과지향적 지표의 비중이 높은 순서



a. <그림 3>의 순서는 성과지표가 객관적이고, 결과지향적 지표(RO)인 비중의 순서이다. 성과지표에 따라 객관적 지표이나 작업량 지향적 지표(WO)인 경우도 있으며, 주관적 지표이나 결과지향적 지표(RS)인 경우도 있지만 혼합값들은 계산에서 제외하였다. 이는 성과지표가 측정가능하고(객관적 지표), 유의미(결과지향적)하게 설정되었는가를 명확하게 판별하기 위해서이다.

<그림 3>은 성과지표가 객관적 지표 및 결과지향적 지표로 구성되어 있는가 여부를 평정한 후 그 비중이 높은 기관 순으로 나열한 것이다. 그 결과 관세청, 해양경찰청, 국토해양부, 기상청 등이 성과지표가 객관적이고 결과지향적 지표(RO)로 구성된 비중이 매우 높았다. 반대로, 기획재정부, 여성부, 통일부, 외교통상부 등은 성과지표가 주관적이고, 작업량 지향적인 성격이 매우 높다고 할 수 있다.

한편, 변수 간 상관관계도 주목할 만하다. 지표가 객관적(O)이고 결과지향적(R)일 때, 단위 사업이 유형적 단위 사업으로 평정된 상관계수가 0.78로 매우 높게 나타났다(부 평정자의 경우 0.67). 이는 유형적 단위 사업인 경우 그 성과지표가 객관적이고 결과지향적일 가능성이 매우 높다는 것을 의미한다. 반대로 무형적 단위 사업일인 경우 성과지표가 상대적으로 주관적이고, 작업량 지향적으로 구성되었을 가능성이 높다는 것으로 유추할 수 있다. 따라서 단위 사업의 유형성과 성과지표의 성격은 매우 밀접한 관계를 가지고 있다고 볼 수 있다 (<표 8> 참조).

<표 8> 성과지표와 단위 사업의 유무형성간의 상관관계

	RO_rat~G	WS_rat~G	TAN_ra~G	RO_rat~Y	WS_rat~Y	TAN_ra~Y
RO_ratioG	1.0000					
WS_ratioG	-0.5128*	1.0000				
TAN_ratioG	0.7187*	-0.4197*	1.0000			
RO_ratioY	0.9191*	-0.5714*	0.6721*	1.0000		
WS_ratioY	-0.4978*	0.5183*	-0.3706*	-0.5053*	1.0000	
TAN_ratioY	0.7497*	-0.4923*	0.8255*	0.7872*	-0.5052*	1.0000

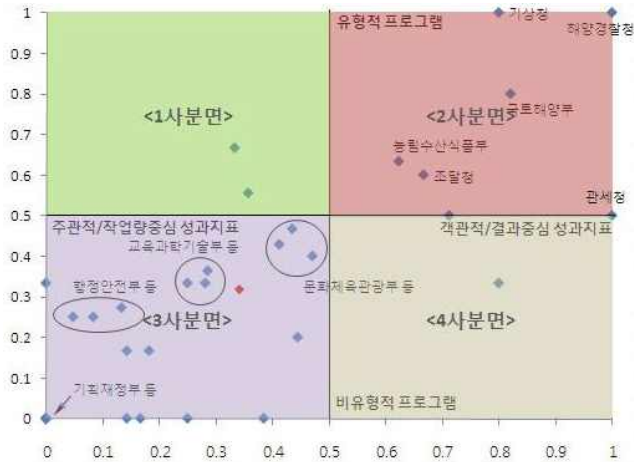
* P<0.05

- a. RO_rat~Y : 주 평정자의 객관적/결과지향적 성과지표 평정값
- b. WS_rat~Y : 주 평정자의 주관적/작업량 지향적 성과지표 평정값
- c. Tan_rat~Y : 주 평정자의 유형적 사업 평정값
- d. RO_rat~G : 부 평정자의 객관적/결과지향적 성과지표 평정값
- e. WS_rat~G : 부 평정자의 주관적/작업량 지향적 성과지표 평정값
- f. Tan_rat~G : 부 평정자의 유형적 사업 평정값

4. 성과주의 예산제도와 기관별 성과측정 적합도

이상에서 성과지표의 객관적 지표 및 결과지향적 지표 여부와 단위 사업의 유무형성 여부를 기관별로 평정하였다. <그림 4>은 각 기관별 평정 결과를 사분면에 표시한 것이다. 성과주의 예산제도의 내적 변수라고 할 수 있는 성과지표의 측정 가능성, 사업의 특성을 고려하였을 때 <그림 4>에 나타난 부처들의 분포를 통해 어떠한 기관이 성과주의 예산제도와 성과관리에 적합한지를 유추해 볼 수 있다.

<그림 4> 단위 사업 및 성과지표 분석을 통한 기관들의 산점도(scatter plot)



<그림 4>에 따르면 국토해양부, 농림수산식품부, 조달청 등이 2사분면에 위치한다. 이들 기관의 단위 사업의 유형성이 높고, 성과지표 역시 객관적이고, 결과지향적 지표로 구성되었음을 보여준다. 따라서 앞서 논의된 성과주의 예산제도 또는 성과관리의 특성을 고려할 때 매우 적합한 기관들로 평정되었다.

또한, 기획재정부, 행정안전부, 통일부, 교육과학기술부 등은 3사분면에 위치하여 성과주의 예산제도와와의 적합정도가 낮은 것으로 평정되었다. 특히 33개의 행정기관 중 23개의 기관이 3사분면에 위치하여 단위 사업의 유형성이 낮고, 성과지표의 객관성, 결과지향성이 낮은 것으로 평가되었다. 이는 이들 기관이 수행하는 사업의 특성이 성과관리에 적합하지 않은 무형적 사업의 비중이 높다는 것을 보여주는 것이다⁷⁾. 또한 성과지표의 경우 주관적 지표의 비중과 단순한 산출물(output)의 측정을 위한 성과지표의 비중이 높은 것으로 나타났다.

한편 1사분면에 위치한 국방부와 산림청의 경우 단위 사업의 유형성이 높았음에도 불구하고, 성과지표의 객관성과 결과지향성 낮은 것으로 평가되었다. 이는 성과지표가 단순히 산출물(output) 측정에 그쳤기 때문이다. 이들 기관은 특별히 성과지표의 재설정이 필요한 것으로 보인다. <표 8>에서 보는 바와 같이 사업의 유형성과 성과지표의 객관성/결과지향성은 높은 상관관계를 나타내는데, 사업의 유형성이 높음에도 불구하고 객관적, 결과지향적 지표의 비중이 낮은 것은 결국 성과지표의 설정을 사업의 특성을 반영하지 못한 채 주관적으로 설정

7) 매년 재정사업의 1/3을 평가하는 재정사업 자율평가의 특성상 당해 연도(2008년도)에만 유독 유형성이 낮은 단위 사업이 평가대상으로 선정되었을 가능성을 배제할 수는 없다. 이를 보완하기 위해서는 이론적 논의와 결과와는 배치되는 결과가 나온 기관의 경우 3개년치의 단위 사업을 모두 분석하는 등의 심층적 분석이 필요할 수 있다.

하였거나 단순한 작업량을 측정하는 지표를 설정하였기 때문이다.

2사분면에 위치한 관세청, 농림수산물식품부, 지식경제부, 국토해양부, 기상청, 조달청, 해양경찰청 등은 모두 단위 사업이 유형적인 동시에, 성과지표 역시 객관적이며, 결과지향적인 성격을 가진 것으로 평가된 기관이다. 앞서 밝혔듯이 단위 사업의 유형성과 성과지표의 객관성 및 결과지향성은 상당한 상관관계를 가지고 있다. 그러나 단위 사업이 측정이 용이한 유형적 성격임에도 불구하고, 성과지표가 성과목표와 연계성이 떨어지는 기관들도 존재하기 때문에 객관적이고, 결과지향적 성과지표를 개발한 해당 기관의 노력은 평가받아야 한다. 즉, 기관 별로 성과관리에 대한 이해도와 숙지도가 다른 현상은 해결되어야 할 과제라고 볼 수 있다.

4사분면에 위치한 식품의약품안전청과 2사분면과 4사분면을 구분하는 X축 위에 위치한 지식경제부, 관세청은 별도의 해석이 필요하다. 식품의약품안전청의 경우 성과지표는 객관적이고 결과지향적으로 설정되었는데, 수행하는 사업의 성격은 무형적이다. 이는 사업의 성격에 비해 성과지표의 개발이 우수한 기관으로 판단될 수 있다. 지식경제부, 관세청 역시 사업 성격에 비해 성과지표의 객관성, 결과지향성이 높은 것으로 조사되었다.

5. 기관별 성과측정 적합도(PMA) 산출

성과주의 예산제도는 앞서 논의한 바와 같이 성과관리 그 자체 뿐만 아니라 한 국가의 행재정적 특성, 역사적 맥락, 정치적 환경, 조직적 특성에 영향을 받을 수밖에 없다. 그러나 우리나라 성과주의 예산제도의 중요한 핵심 축인 재정사업 자율평가 제도 분석을 통해서 특정 시점의 각 기관별 성과측정 적합도(PMA)를 측정하는 것은 가능할 것이다. 성과측정 적합도는 각 기관의 단위 사업의 특성(단위 사업의 유형성 여부)은 어떠하며, 이에 부합하는 적절한 성과목표와 성과지표를 설정하고 있는가를 나타내는 지표이다. 성과측정 적합도는 각 기관의 성과주의 예산제도와와의 보편적인 적합 정도를 의미하는 것은 아니다. 특정 시점의 적합 정도를 뜻하며, 성과측정 적합도는 제도의 지속적 시행으로 인한 성숙, 조직 학습(organizational learning)⁸⁾ 등으로 개선될 여지가 있다. 실제로

8) 조직목표의 효과적 달성을 위해 과거의 경험에 비추어 특정한 행동들을 강화하거나 수정하면서 환경 변화에 적응해 가는 조직의 학습활동을 말한다. Argyris(1977)는 조직의 유지·성장에 기여했던 행동 및 대응 양식은 반복되고 그렇지 못한 행동들은 중단되거나 수정되어야 한다고 하면서, 이 때 행동 및 대응 양식의 수정만을 통해 조직 목표를 달성하려는 활동을 1차 학습 또는 적응학습이라 하고, 조직의 목표 자체까지도 수정하는 학습을 2차 학습 또는 생성학습이라고 했다. 1차 학습은 일상적이고 반복적인 업무에 적절하고, 2차 학습은 복잡하고 프로그래밍되지 않은 업무의 학습에 적절하다.

미국 GAO(1999)에 따르면 각 연방기관들이 PART⁹⁾ 시행 초기보다 성과관리, 성과지표 설정 등에서 높은 이해도를 보이는 것으로 조사되었다. 우리나라의 PART라고 할 수 있는 재정사업 자율평가 역시 시범기간을 포함한다면 10년이 경과한 시점이기 때문에 초기보다 단위 사업, 성과목표, 성과지표 등 성과정보의 제시 수준은 분명히 향상되었을 것이다. 향후 제도의 시행이 경과한 이후 다시 성과측정 적합도를 측정한다면 의미 있는 결과와 정책적 시사점을 도출해 낼 수 있을 것이다. 성과측정 적합도는 각 기관들이 자체 성과를 관리하기에 얼마나 적합한지, 또 적절한 지표를 사용하고 있는지를 나타내는 것이다. 따라서 실제로 성과관리를 적절히 하고 있는지 여부는 전혀 별개의 문제이다.

<표 9>는 앞 장에서 분석한 33개 기관들의 단위 사업, 성과지표 분석을 통해 계산된 성과측정 적합도 수치이다. 수식은 다음과 같이 나타낼 수 있다.

$$PMA^* = \left(\frac{\text{객관적} \cdot \text{결과지향적 성과지표의 수}}{\text{성과지표의 수}} + \frac{\text{유형적 프로그램의 수}}{\text{프로그램의 수}} \right) \times 0.5$$

*PMA : performance measurement appropriateness(성과측정 적합도)

- a. 성과측정 적합도를 나타내는 위 공식에서 성과지표 값과 프로그램 값에 각각 0.5를 곱한 것은 PMA의 총점을 1점 만점으로 나타내기 위해서임. 그런데, 성과지표 값과 프로그램 값의 가중치가 다를 경우 차별된 값을 곱할 수 있을 것임. 본 연구에서는 두가지 척도, 즉 객관적·결과지향적 성과지표 비중과 유형적 프로그램 비중을 동일하게 적용하였기 때문에 동일하게 0.5를 곱하였음.

성과측정 적합도를 산출해본 결과 해양경찰청, 기상청, 국토해양부, 관세청 등이 최상위권을 차지하였다. 이는 이들 기관의 단위 사업이 매우 유형적인 동시에 단위 사업 성과목표에 의해 설정한 성과지표 역시 객관적이고, 결과지향적인 성격이 강했다는 의미이다. 그런데 주목할 만한 사실은 성과측정 적합도가 높은 이들 4개 기관 중 국토해양부를 제외하고는 모두 기관의 규모가 작은 ‘청’ 단위 기관이었다. 이는 조직의 목표가 상대적으로 간명하고, 조직의 규모가 상대적으로 작기 때문에 단위 사업과 성과지표의 모호성이 축소될 수 있다. 또한 청 단위를 제외한 부 수준에서 기획재정부, 외교통상부, 통일부가 성과측정 적합도가 가장 낮은 기관으로 조사되었다. 이들 부처의 사업의 특성, 성과지표의

9) PART(Program Assessment Rating Tool) : 사업성과를 평가하는 체계적 도구이면서 사업 성과 향상을 위한 진단적 도구이고, 예산편성을 위한 정보 제공을 위한 제도이다. PART는 평가항목을 사업의 목적 및 설계, 전략적 기획, 사업관리, 사업 성과와 책임성 등 네 가지로 분류하고, 각 평가항목 당 질문 문항은 5-10개로 총 25개로 구성되어 있다.(OMB, 2005) 우리나라의 재정사업 자율평가제도가 미국의 PART를 벤치마킹한 것이다.

객관성 및 결과지향성이 매우 낮은 것이다.

〈표 9〉 기관별 성과측정 적합도

순위	기관명	성과측정 적합도
1	해양경찰청	1.00
2	기상청	0.90
3	국토해양부	0.81
4	관세청	0.75
5	조달청	0.63
6	농림수산식품부	0.63
7	지식경제부	0.61
8	식품의약품안전청	0.57
9	산림청	0.50
10	국방부	0.46
11	환경부	0.45
12	특허청	0.44
13	문화체육관광부	0.42
14	교육과학기술부	0.32
15	중소기업청	0.32
16	보건복지가족부	0.31
17	국세청	0.29
18	문화재청	0.20
19	노동부	0.19
20	법무부	0.17
21	방송통신위원회	0.17
22	여성부	0.17
23	통계청	0.15
24	행정안전부	0.15
25	농촌진흥청	0.13
26	경찰청	0.08
27	국가보훈처	0.07
28	금융위원회	0.00
29	기획재정부	0.00
30	병무청	0.00
31	소방방재청	0.00
32	외교통상부	0.00
33	통일부	0.00

- a. 성과측정 적합도가 1점 만점을 기록한 해양경찰청은 모든 사업이 유형적인 성격을 가지고 있으며, 성과지표 역시 결과지향적이고, 객관적 지표로 구성되어 있음을 뜻함. 반대로 성과측정 적합도가 0점을 기록한 금융위원회, 기획재정부, 병무청, 소방방재청, 외교통상부, 통일부 등은 모든 사업이 무형적이면서, 성과지표 역시 작업량 중심적이고, 주관적 지표로 구성되어 있음을 의미함. 그러나 성과측정 적합도 지수만으로 성과측정의 정도를 판단하기에는 무리가 있다. 특히 성과측정 적합도 지수가 '0'인 기관 들 중 지수 측정 대상 사업의 개수가 현저히 적은 경우 별도의 분석이 수반되어야 할 것이다.

또한, 이론적 논의를 기반으로 한 선행연구에서 시도했던 정책의 유형 분류가 <표 9>를 통해서 대체로 일치하고 있다고 판단할 수 있다. 그러나, 현대 국가에서 정부의 책무가 다양해지고, 복잡성이 가중되면서 유형성이 분명하다고 관념적으로 판단할 수 있었던 기관들도 무형적 프로그램을 시도하는 것이 관찰되었다. 반대로 무형성이 강하다고 인식되었던 기관들이 유형적 프로그램을 시도하는 등 기관에 대한 심층 분석이 수행되지 않으면 유무형성을 단정적으로 논하기 힘들다고 판단된다. 또한 국가적 특수성으로 인해 기관들의 성격을 각 국가마다 수평적으로 비교 분석하기 어렵고, 기능 역시 수시로 변화한다. 뿐만 아니라 현대 국가는 다양한 정책 수단을 동원하고 정책의 혼합을 통해 그 효과를 제고하기 위한 노력을 하기 때문에 정책 유형을 분류하기가 관념적으로 판단하는 것만큼 쉬운 일은 아니다. 따라서 성과측정 적합도 같은 개념을 통해 수시로 해당 기관의 정책적 성격과 성과지표의 적절성 등을 파악하는 것은 매우 유의미한 일이라고 할 수 있다.

IV. 결 론

이상에서 기관별 재정사업 자율평가 최종보고서에 기재된 단위 사업과 성과지표의 성격을 분석하였다. 분석 방법은 33개 기관이 제출한 최종보고서의 단위 사업과 성과지표를 모두 수집하여 성과주의 예산제도와 관련된 주요 쟁점 중 외생변수를 제거하고 추출하였던 단위 사업의 유무형 여부와 성과지표의 분류를 시도하였다. 또한 단위 사업의 유무형성, 성과지표 분류의 타당성(validity)의 증거를 확보하기 위해 주연구자-부연구자 사이의 평정자간 신뢰도(inter-rater reliability)를 측정하여 상관관계 계수를 확보하였다.

그 결과는 다음과 같이 요약할 수 있다.

첫째, 선행연구에서 시도하였던 포괄적인 정책 영역 구분을 우리나라 재정사업 평가 보고서에 기재된 단위 사업 및 성과지표 분석을 통해 실증적으로 검증하였다. 그 결과 각 기관별 단위 사업의 유무형성과 성과지표의 적절성(객관성, 결과지향성)의 정도를 도출해내었다. 단위 사업의 유형성(tangibility)이 높은 기관은 기상청, 해양경찰청, 국토해양부, 산림청 등이며, 단위 사업이 유형성이 낮은 기관은 금융위원회, 국가보훈처, 노동부, 외교통상부, 통일부 등이었다. 또한 관세청, 해양경찰청, 국토해양부, 기상청 등이 성과지표가 객관적이고 결과지향적 지표(OR)로 구성된 비중이 매우 높았으며, 기획재정부, 여성부, 통일부, 외교통상부 등은 성과지표가 주관적이고, 작업량 지향적인 성격이 높은 편이었다.

셋째, 변수 간 상관관계는 매우 주목할 만하다. 지표가 객관적(O)이고 결과지

향적(R)일 때, 단위 사업이 유형적 단위 사업으로 평정된 상관계수가 매우 높게 나타났다.(상관계수 0.78) 이는 유형적 단위 사업인 경우 그 성과지표가 객관적이고 결과지향적일 가능성이 매우 높다는 것을 의미한다. 반대로 무형적 단위 사업일 경우 성과지표가 상대적으로 주관적이고, 작업량 지향적으로 구성되었을 가능성이 높다는 것으로 유추할 수 있다. 따라서 단위 사업의 유형성과 성과지표의 성격은 매우 밀접한 관계를 가지고 있다고 볼 수 있다. 이는 단위 사업의 성격이 곧 기관의 성격을 대변한다고 봤을 때, 기관별로 성과주의 예산제도와 적합정도는 상이하다는 증거가 될 수 있다.

넷째, 성과측정 적합도(PMA)의 측정을 통해 현 시점의 각 기관별 단위 사업의 특성과 성과지표 설정의 적절성을 파악할 수 있다. 성과측정 적합도는 각 기관의 성과주의 예산제도와외의 보편적인 적합 정도를 의미하는 것은 아니고, 특정 시점의 적합 정도를 뜻한다. 우리나라 기관들을 대상으로 성과측정 적합도를 산출해본 결과 성과측정 적합도가 높은 이들 4개 기관 중 국토해양부를 제외하고는 모두 기관의 규모가 작은 ‘청’단위 기관이었다. 이는 조직의 목표가 상대적으로 간명하고, 조직의 규모가 상대적으로 작기 때문에 프로그램과 성과지표의 모호성이 축소될 수 있음을 증거하고 있다. 한편, 선행연구에서 시도했던 정책의 유형 분류가 대체적으로 일치하고 있다고 판단할 수 있지만, 현대 국가는 다양한 정책 수단을 동원하고 정책의 혼합을 통해 그 효과를 제고하기 위한 노력을 하기 때문에 정책 유형을 분류하기가 관념적으로 판단하는 것만큼 쉬운 일은 아니다. 따라서 성과측정 적합도 같은 개념을 통해 수시로 해당 기관의 정책적 성격과 성과지표의 적절성 등을 파악하는 것은 매우 유의미한 작업이 될 것으로 전망된다.

이 연구를 통해 도출된 재정사업 자율평가제도(우리나라의 성과주의 예산제도에 대한 정책적 시사점과 개선방안을 정리하여 보면 다음과 같다.

첫째, 우리나라 기관들의 재정사업 자율평가 결과보고서를 분석한 결과 재정사업 자율평가 제도에 대한 기관별 이해도가 상이하고, 이에 따라 성과지표의 타당성을 저해하는 문제가 발생하였다. 단위 사업에 따라서 성과지표의 수가 다를 수밖에 없지만, 측정가능하나 의미가 떨어지는(작업량 지향적) 지표의 사용과 평가 방법이 모호한 지표의 사용은 향후 개선될 필요가 있다. 또한 분석대상 기관 중에는 성과목표와 성과지표를 동일하게 제시하여 성과지표의 기본적 개념에 대한 이해도가 떨어지는 기관도 존재하였다.

둘째, 성과지표 평가방법의 획일적 경향이 목격되었다. 특히, 기관들이 선호하고 있는 설문조사방식은 대체로 평가에 적합한 형식논리에 적합한 형태를 취하고는 있으나, 평가의 핵심인 성과에 대한 평가나 평가 목적 또는 성과목적과는 어울리지 못하는 경향이 있다. 이는 적합하고 다양한 평가방법을 개발하지

못하였기 때문이기도 하고, 손쉬운 성과지표 측정방법을 찾겠다는 각 기관들의 유인이 결합된 현상으로 판단된다. 또한 더 큰 문제는 평가방법에 있어서 기관들에 공통적으로 적용될 수 있는 평가설계모형이 부재함에 따라서 평가기준과 항목 설정에 있어서 자의성과 측정의 오류를 초래하게 될 가능성이 크다는 것이다. 따라서 이론적, 제도적 보완을 통해 평가방법 별로 체계성과 통일성을 확보하는 것이 필요하다고 하겠다.

셋째, 거시적으로 정책평가방법이 타당성을 확보하기 위해서는 통합적인 평가방법체계를 제시하고 운영할 필요가 있다. 또한 국무총리실이 주관하는 정부정책평가 등의 제도와 연계성을 높이고, 정부 평가제도 전반을 간명화할 필요가 있다. 또한 기관 입장에서는 평가 방법이 용이한 방안을 선택할 유인이 있기 때문에 객관적 지표를 제시하는 것뿐만 아니라 정성적 지표에 대한 평가 방안을 도입하여야 한다. 객관적, 과학적 평가 방법을 강조하다 보면 측정 용이한 방법을 선택하게 되는 현상을 방지하기 위해 정량적 측정 방법과 정성적 측정 방법을 결합한 다중방법론의 채택과 정책평가방법 활용의 획일성 문제를 해결하기 위해 다양한 평가 방법을 개발하여야 한다(제갈돈 외, 2006).

참고문헌

- 각 부처. 2009. “2008년도 재정사업 자율평가 보고서”. 기획재정부 자료실.
 기획재정부. 2009. “2008년도 재정사업 자율평가 결과”. 기획재정부 보도자료. 2009년 7월 29일.
 박노옥·원종학·김진·박명호. 2008. “재정성과 관리제도의 현황과 과제”. 한국조세연구원.
 전영한. 2004. “공공조직의 목표 모호성: 개념, 측정 그리고 타당화”. 한국행정학보, 38(5): 49-65.
 제갈돈·김주환·김홍률·이윤식. 2006. “성과관리를 위한 우리나라 평가지표체계의 개선방안에 관한 연구”. 한국거버넌스학회보, 13(2): 217-248.
 조수현. 2007. “예산개혁에 대한 제도주의적 분석”. 연세대학교 대학원 박사학위논문.
 Anderson, James. 1975. "Public Policy-Making". New York: Praeger.
 Argyris, Chris. 1977. "Organizational Learning and Management Information Systems". *Accounting, Organizations and Society*, 2(2): 113-123.
 Bouckaert, G. & Ulens. W. 1998. *Mesure de la performance dans le service public: exemples étrangers pour les pouvoirs publics Belges*. Bruxelles,

Services fédéraux des affaires scientifiques, techniques et culturelles.

- Chun, Young Han & Hal G. Rainey. 2005. "Goal Ambiguity in U.S. Federal Agencies". *Journal of Public Administration Research and Theory*. 15(1): 1-30.
- Grizzle, G. A. 1982. "Measuring State and Local Government Performance: Issues to Resolve Before Implementing a Performance Measurement System". *State and Local Government Review*, 14: 132-136.
- Krippendorff, K. 2004. *Content Analysis: An Introduction to Its Methodology*. Beverly Hills, CA: Sage.
- Lee, Jung wook, Hal G. Rainey & Young Han Chun. 2009. "Of Politics and Purpose : Political Salience and Goal Ambiguity of us Federal Agencies". *Public Administration* 87(3): 457-484.
- Lombard, Matthew, Snyder-Duch, Jennifer & Bracken C. Cheryl. 2002. "Content Analysis in Mass Communication-Assessment and Reporting of Intercoder Reliability". *Human Communication Research*, 28(4): 587-604.
- Pollitt, Christopher. 1999, "Integrating Financial Management & Performance Management". *OECD Journal on Budgeting*, 1(2): 7-37.
- Robinson, Marc. 2007. *Performance Budgeting: Linking Funding and Results*. IMF.
- Sabatier, Paul A. 1991. "Toward Better Theories of the Policy Process". *Political Science and Politics*, 24(2): 147-156.
- Salamon, Lester M. 2002. "The New Governance and the Tools of Public Action: An Introduction". In Salamon, Lester M. (ed.) *The Tools of Government: A Guide to the New Governance*. Oxford University Press.
- Vedung, Evert, Marie-Louise Bemelmans-Videc, & Ray C. Rist, 2003. *Carrots, Sticks, and Sermons: Policy Instruments and Their Evaluation*. New Brunswick, New Jersey: Transaction Publishers

Abstract

An Analysis of the Appropriateness of Performance Budgeting by Government Departments: The Development & Application of "Performance Measurement Appropriateness"

Youngrok Kim

This study aims to conduct an analysis of Korea's case by designating policy types and performance indicator traits as variables made up of the key issues proposed in previous studies regarding performance budgeting. The study identifies relevant ministries and policies subject to performance budgeting after appraising the suitability of the performance indicator and classifying the policy types at all of the ministries which employ the Performance Assessment Rating Tool (PART) in Korea. A scale of 'Performance Measurement Appropriateness' was also developed for common use based on a suitability analysis of performance indicators set for unit projects in the execution process. As a result, a discriminative analysis of the performance measurement appropriateness of Korea's central administrative agencies was worked out.

【Key words: Performance budgeting, performance measurement appropriateness, Program Assessment Rating Tool (PART), policy sector, tangibility, performance indicator】