

선진국 공공연구개발프로그램 평가시스템의 비교분석

Comparative Analysis on the Evaluation Systems of the Public R&D Programs in the Developed Countries

홍 형 득*

〈目 次〉

- | | |
|-----------------------|-----------------|
| I. 서 론 | IV. 평가시스템 비교 |
| II. 평가시스템 구성요소 및 연구모형 | V. 정책적 시사점 및 한계 |
| III. 프로그램의 개요 및 구조 | |

<Abstract>

The frameworks for evaluation of national R&D programs reflect their various political and administrative cultures(Gibbons & Georghiou, 1987), and the structure of national R&D system. In this research, the core research questions are: what is good evaluation? What is being evaluated, by whom, by which criteria, for whom and what purpose?

In order to examine these general aims and answer these questions, in detail, several objectives can be proposed on the process of this research. In this research, the national R&D programs will be considered in terms of the interface between evaluation and the wider policy-making process. The programs for case study are the Alvey program(UK), the Advanced Technology Program(US) and the Framework program(EU). One of the characteristics of these programs is the well established evaluation systems they have. From the comparative analysis, we can withdraw some useful implication for Korean evaluation practice for national R&D programs. Evaluation system is social process and the way in which it is organized is critical to its outcome.

Key words: 연구개발 프로그램, 평가시스템, 평가결과

*국립 밀양대학교 행정학과 조교수, hdhong@arang.miryang.ac.kr

I. 서론

최근 연구개발프로그램들이 복합화, 대형화되고 있고 특히 WTO체제의 등장과 아울러 과학기술에 대한 국가적인 관심의 증대와 국가간의 기술경쟁이 가속화되면서 선진국을 중심으로 한 많은 국가에서 정부와 산·학·연 모든 연구개발주체가 총동원되는 국가주도의 대형연구개발프로그램들이 중요한 과학기술정책 수단으로 등장하였다. 이처럼 1990년대 이후 과학기술정책의 중요한 특징 중의 하나는 이런 공공프로그램을 통한 정책적인 연구개발지원 및 육성전략이다. 이들 대형연구개발프로그램들은 장기간 대규모의 인력과 자원이 투자되기 때문에 전체적인 프로그램 관리가 어렵고, 연구개발의 실패가능성과 실패할 경우 국가경제적 충격이 크기 때문에 상황변화에 따라 신속한 전략의 수정이 필요하고, 재정적인 긴축과 과학기술에 대한 국민의 관심과 참여가 증대되어 공공자금 사용의 정당성과 성과에 대한 요구의 증대 등 공공연구개발프로그램의 평가가 그 어느 때보다 중요하게 부각되고 있다 (Georghiou, 1987). 그러나 국가연구개발프로그램은 그 목적이 포괄적이고, 효과가 광범위할 뿐만 아니라 장기간이 지나서야 성과측정이 가능하기 때문에 평가에 그만큼 어려움이 있다 (Roessner, 1989). 우리나라에서도 1982년 과학기술부의 특정연구개발사업을 시작으로 대부분의 부처에서 연구개발사업을 개발해오고 있으며, 성과주의예산의 점진적인 도입과 아울러 평가의 중요성과 그 역할이 커지고 있다.

따라서 본 연구는 국가대형연구개발프로그램을 위한 적합한 평가시스템(Good Evaluation System)을 찾기 위한 시도이며, 선진국의 대표적인 국가전

략 연구개발프로그램들의 평가시스템을 비교해 봄으로써 국가간 환경 및 프로그램의 특성에 따른 평가시스템상의 특징들을 파악하고 한국에의 시사점을 얻고자 하는 것이다. 본 연구의 연구대상 프로그램은 영국의 Alvey, 미국의 ATP(Advanced Technology Program) 그리고 유럽연합의 Framework프로그램이며, 이들의 공통적인 특징은 국가의 전략적인 연구개발프로그램으로 국가적인 관심의 초점이 되고, 철저한 프로그램 평가시스템을 갖추고 있다는 점이다.

본 연구를 위한 분석자료는 각 프로그램 평가보고서들에 나타난 사실을 중심으로 한 문헌조사를 주로 하고 평가책임자와 프로그램 관리자들과의 면접조사방법(Interview)을 통해 보완하였다¹⁾. 면접조사는 1999년 12월부터 2000년 3월까지 이루어졌다. 면접조사 대상자와 사전약속을 하고, Alvey프로그램을 위한 영국 맨체스터대학교의 PREST와 ATP 프로그램을 위한 미국의 NIST, 그리고 Framework 프로그램을 위하여 벨기에의 유럽연합 본부(European Commission) 등을 방문하여 1-2시간의 일대일(Face to Face) 면접조사가 이루어졌다. 평가책임자와 프로그램 관리자들과의 면접방법은 반구조적 설문(Semi-structured questionnaire)을 이용하여 각 프로그램의 평가시스템에 관한 자유로운 의견을 들을 수 있도록 설계하였다.

II. 평가시스템 구성요소 및 연구모형

평가시스템은 하나의 사회적인 과정이라고 할 수

1) 각 프로그램별 인터뷰 대상자는 다음과 같다.

Alvey : Georghiou(Univ. of Manchester), Cameron(Univ. of Manchester), Hill(DTI)

ATP : Reugg(NIST), Powell(NIST), Spivack(NIST), Etzkowitz(State Univ. of New York)

Framework : Fayl(Head of Evaluation Unit, EC), Hartley(Head of Unit, EC), O' Sullivan(Evaluation Unit, EC) 외 다수

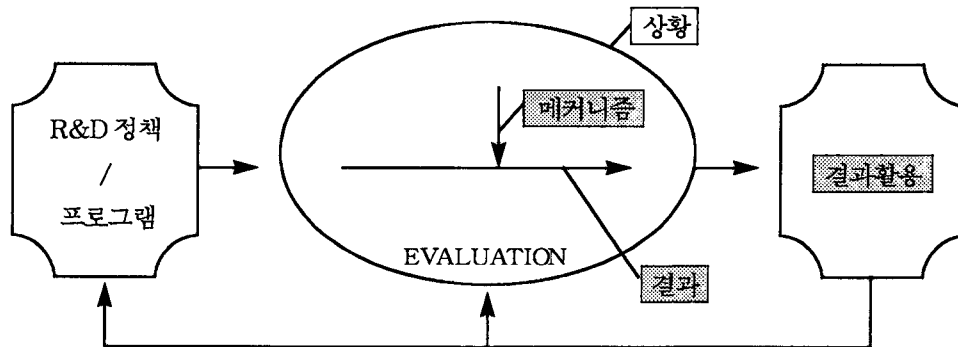
있으며, 따라서 국가에 따라 그리고 그 평가의 대상에 따라 평가시스템은 다르고, 그만큼 다양하다. 각 프로그램 평가시스템을 체계적으로 비교분석하기 위해서는 구체적인 구성요소들이 결정되어야 하는데, 연구개발사업의 평가시에 나타나는 대표적인 이슈들은 누가, 무슨 기준으로 누구를 위하여 그리고 무슨 목적으로 평가하는가 등이다(Georghiou, 1989).

Scriven(1991)은 '정책평가의 8가지 접근방법'을 통해 평가와 관련된 기본적인 문제로 목적, 평가자, 개입, 집행, 결과, 영향, 기준, 활용 등의 8가지 구성요소들을 제시하였다. Vedung(1991)은 이들을 다시 개입의 평가, 개입 그 자체, 그리고 피드백과 활용이라는 세가지로 재구분하였다. 특히 연구개발사업의 평가에 관하여 Georghiou(1989)는 평가범위, 조직화, 평가기준 및 평가목적의 4가지를 설정하였고,旭 Research Center(1988)는 평가자, 평가방법, 평가 형태를 포함하는 좁은 의미의 평가와 연구평가의

지원시스템을 포함하는 넓은 의미의 평가시스템 개념을 제시하고 있다. 이처럼 평가에 대한 기본적인 관점은 동일하다 하더라도 평가대상에 따라 그 구성요소나 개념들이 다를 수 있다.

지금까지의 평가시스템 구성요소(Components)들은 <그림 1>에서 보는 바와 같이 평가의 목적 및 시기와 평가자, 및 평가대상을 포괄하는 평가상황(Context), 평가항목, 평가방법, 의사소통, 평가지원요소 등 과정요소를 포함하는 평가메카니즘(Mechanism), 평가결과(Outcome), 그리고 그 평가결과의 활용(Utilization) 등으로 구분할 수 있다. 각각의 가이드라인들의 관점이나 기준들이 다양하고 실제로 그 대상이 다르기 때문에 이들을 종합한 평가의 원칙을 만든다는 것은 쉽지 않다. 따라서 본 연구에서는, 평가시스템의 네 가지 범주에 기존 연구들에 나타난 각 구성요소들의 항목을 재구분하여 <표 1>과 같이 구성하였다.

<그림 1> 메타평가를 위한 평가시스템의 기본구조 및 모형



<표 1> 평가시스템의 범주 및 구성요소

범 주	구 성 요 소	
I. 평가상황	1. 평가배경 3. 평가자	2. 평가관련자 4. 평가비용
II. 메카니즘	5. 평가기준 7. 커뮤니케이션 메카니즘	6. 평가방법론
III. 평가결과	8. 보고서 및 정책제안	9. 결과보고
IV. 결과활용	10. 활용을 위한 전략	11. 활용유형

1. 평가상황(Context)

평가의 상황은 대체로 평가의 배경, 평가자, 평가관련자(Stakeholders), 그리고 평가 자원 등 평가의 기본적인 요소들이다. 여기에서는 그 프로그램 평가를 둘러싼 정치적, 사회적 환경 및 평가의 목적 등이 파악되어야 한다. 그 프로그램과 프로그램이 갖고 있는 상황에 대한 명확한 파악이 필요하고, 평가의 목적 및 과정에 대해 구체적으로 논의되어야 한다.

평가관련자에 관하여는 평가관련자들의 관심사항 파악 및 정보제공이 중요하고 이와 관련하여 어디로부터 평가가 요구되었는지도 평가의 과정과 특성에 중요한 영향을 미치기 때문에 고려되어야 할 사항이다. 가장 중요한 요소 중의 하나인 평가자와 관련한 이슈는 우선 평가자의 능력과 신뢰성 및 독립성이 고려되어야 하며, 평가자 선정과 평가가 평가자에게 어떻게 의뢰되는지의 계약형태 등이 중요한 요소들이다. Weiss(1998)는 내부참모, 외부평가기관, RFP 및 대학 연구지원 등 4가지 유형으로 구분하고 있다. 평가비용은 항상 논란의 대상이 되지만 일반적인 원칙은 없다. 다만 평가비용과 관련한 몇 가지 논의는 거시적 차원(Macro-level)에서 전체 프로그램예산의 몇 퍼센트가 평가비용으로 적절한 것인가? 혹은 미시적 차원(Micro-level)에서 평가 예산내에서 각 예산의 배분구조는 어떻게 되는가? 등이다(Alkin & Stecher, 1983). Drezek et al.(1982)는 조사결과 평가비용이 전체프로그램 예산의 1.5%에서 5.5% 사이에 있었다고 했고, Owen & Rogers(1999)는 10-15%를 제안하는 등 많은 차이가 있다. 본 연구에서의 평가비용은 거시적 차원에서만 논의가 이루어질 것이다.

2. 집행 메카니즘(Mechanism)

평가과정은 프로그램의 가치와 장점(Value and Merit)을 판단하기 위하여 정보를 수집하고, 사용하

는 과정에서의 여러 활동 메카니즘에 분석의 초점이 두어진다. 그 과정은 평가의 기준과 평가자의 평가를 위한 데이터 수집 및 분석방법, 평가자와 평가관련자들간의 의사소통 메카니즘 등을 대상으로 하고 있다.

평가는 그 대상이 무엇이나에 따라 기준이 달라지게 되는데, Georghiou(1995)는 대상을 연구자, 연구과제, 연구사업, 연구기관, 및 정책수단으로 구분하고 있다. 개인수준의 연구는 조직내에 있는 연구자의 실적이나 역량의 평가이고(Ahmad, 1981), 프로젝트 수준에서는 조직에서 행한 연구개발 프로젝트의 실적 및 공헌도를 평가하는 것이며, 프로그램 수준에서는 역시 개별 프로젝트의 평가를 포함한 프로그램 전체의 성과, 추진체제 및 전략을 평가하는 것이다(Frame, 1983; Meyer-Kraymer, 1984; Schainblantt, 1982).

평가기준은 평가에서의 판단의 근거가 되는 것이며, 평가목적과 부합하는 평가요인이다. 대부분의 평가시스템에는 하나이상의 평가기준을 포함하고 있고, 연구단계별, 평가시기별, 평가대상 등에 따라 차이가 있다. 절대적 평가의 경우 평가의 기준은 과제의 채택 및 기각을 결정하기 위한 판단의 기준이 되기도 하며, 상대적 평가의 경우는 평가항목에 대응하는 특성을 정량화하기 위한 척도가 된다.

평가방법은 평가를 위한 데이터를 수집하고 분석하는 방법으로 정성적인 방법과 정량적인 방법으로 나눌 수 있다. 정성적인 방법으로는 문헌분석, 사례연구, 면접조사 등의 방법을 사용하고, 정량적인 방법으로는 결정론적 평가, 경제론적 평가, OR적 평가, 혹은 선별적 평가방법, 집합적 평가방법 등이 있다. 여기서의 주요 이슈들은 데이터들이 어떻게 선택되고, 어떤 수단을 동원하여 관리되는가? 그리고 데이터와 정보들이 어떻게 수집, 저장 및 기록되는가? 그리고 프로그램의 영향평가를 위하여 어떤 툴(Tool)들이 사용되었는가? 등이다. 평가과정에 있어서 중요한 요소 중의 하나는 평가자와 정책결정자 혹은 프로그램관리자, 그리고 연구자 등을 포함한 평가관련자들과의

커뮤니케이션 메카니즘이다. 이들 커뮤니케이션을 통한 정보교환, 연구방향설정, 평가피드백 등이 이루어지기 때문이다. 의사소통과 관련하여는 평가결과의 한계나 범위에 대하여 관련자들과 충분한 의사소통이 이루어져야 한다.

3. 평가결과(Outcome)

대표적인 평가결과물은 평가보고서와 평가자들이 제시한 정책제안으로 나타난다. 보고서의 형식이나 내용여하에 따라 그 활용여부나 일반에의 이해 등에 영향을 미친다. 따라서 평가결과에 관하여는 평가보고서의 구성과 정책제안사항 그리고 평가보고서의 확산을 위한 평가자들의 보고(Reporting) 등의 노력을 대상으로 한다.

평가결과의 보고에는 바쁜 정책결정자들이 이해하고, 실질정책과정에 활용될 수 있도록 하는 전략이 필요하고, 정책제안은 얼마나 그 제안이 실질적이며 적용가능한지의 여부가 관건이다. 정책제안의 수용 혹은 거절이나 그 제안의 반영과정 등과 평가결과의 보고에서 어휘, 길이, 형식, 시간, 위치 등이 신중히 고려되어야 한다. Scriven(1991)은 최종보고서는 초안의 회람을 통해 사전 스크리닝되어야 하고, 평가보고서는 과학적이고, 상식적인 메시지를 가지고 있어야 하며, 가시적인 데이터나 그림 등 심미적인 자료가 포함되어 있을 것을 제안하고 있다.

4. 결과활용(Utilization)

평가결과는 자동적으로 정책결정과정으로 투입되어 활용되지는 않으며, 활용된다고 하더라도 여러 형태의 활용유형이 있다. 평가결과 활용의 개념은 대체로 좁은 의미와 넓은 의미의 두 가지의 서로 다른 시각이 있다(Weiss, 1977). 좁은 의미의 활용은 평가의 활용을 평가보고서에 제시된 결론이나

건의에 따라 해당 정책이나 사업의 지속적인 실시나 확대, 혹은 중단이나 수정실시 등 즉각적으로 해당 정책이나 사업과 관련된 어떤 결정을 내리는 것이며, 넓은 의미의 활용은 수단적 활용(Instrumental use)과 개념적 활용(Conceptual use), 이 양자를 모두 포함하는 개념으로 본다. 수단적 활용은 평가에서 제시된 결론을 특정한 결정에 적용하는 경우를 의미하고, 개념적 활용이란 평가의 결과가 직접적인 영향을 미치지 않더라도 정책결정자나 사업관리자 등 결과의 이용자들의 정책이나 사업에 대한 이해에 영향을 미치는 것을 의미한다. 평가의 활용 여부는 정도의 문제라고 할 수 있다. 이처럼 평가는 합리적이고 정당한 목적으로 사용되지만 평가의 정치성 등으로 인하여 사업 효과의 가식, 결정의 지연, 책임회피, 대외홍보 및 정책효과 속임 등 바람직하지 못한 비합리적인 목적으로 활용되기도 한다(Weiss, 1972; Suchman, 1972). 평가결과의 활용 부문에서의 주요 이슈들은 평가의 활용을 위한 특정한 전략이 있는가? 평가가 구체적인 의사결정에 어떻게 영향을 미치는지? 정책제안들이 어떻게 정책변화에 영향을 미치는가? 한편 결과의 활용을 위한 메카니즘의 측면에서는 평가결과의 활용메카니즘은 어느 정도 갖추어져 있는가? 얼마나 공개적으로 평가결과에 관해 논의하는가? 결과활용을 위한 커뮤니케이션 채널은? 등으로 요약된다.

III. 프로그램의 개요 및 구조

1. Alvey 프로그램

영국의 Alvey 프로그램은 1983년 첨단정보기술 개발을 위하여 무역산업부(DTI), 국방부(MOD), 과학공학연구심의회(SERC)가 공동으로 참여한 다부처의 전략적인 연구개발프로그램으로 특히 성공적인 실시간 평가(Real Time Evaluation)시스템으

로 잘 알려져 있다(Guy et al., 1991). 공동연구를 원칙으로 하여 관련분야 산학연 연구주체들간의 활발한 교류와 공동 참여를 유도하고, 여러 정부기관, 기업체, 학술기관에서 선정된 인사들로 구성된 사무국(Directorate)이 프로그램의 관리 및 운영을 책임지도록 한 점 등이 프로그램 수행체제상의 특징이다.

초기 프로그램의 주요한 기술분야는 지능형 지식베이스시스템, 소프트웨어 엔지니어링, 초고집적 반도체, 맨머신인터페이스의 네분야였으며, 각 분야에 추진위원회가 구성되었다. 이러한 네가지 기술분야에 대하여 각기 대규모 시범장치를 설치하여 실제 응용을 위한 기반기술을 개발하고, 실제로 프로그램결과를 시행하는 등의 두 가지 목적을 동시에 추구하였으며, 이밖에 연구를 지원하기 위한 통신 및 하부구조에 대한 연구도 수행하였다. 산업계의 참여기업의 경우 예산의 반을 자체조달하고 반은 정부가 지원하도록 하였으며, 참여대학에 대해서는 SERC를 통해 전액 지원하였다. 프로그램의 목표는 5년간 정보기술수준을 2배이상 높이고 일련의 기술목표를 달성함으로써 세계시장에서 영국정보기술산업의 경쟁력 강화를 목표로 하였다.

2. ATP 프로그램

미국의 첨단기술프로그램(ATP)은 상무부(DOC) 산하 국립표준기술원(NIST)에서 관리하고 있는 프로그램으로서, 산업의 기술경쟁력 강화를 위한 미국의 대표적 연구개발프로그램이다²⁾. 전통적으로 민수산업의 연구개발에 대해서는 민간기업의 노력에 맡겨온 미국이 이 프로그램을 추진하게 된 배경에는 1980년대 이후 미국의 산업기술 경쟁력이 일본에 크게 위협받으면서 연방정부의 기술정책이 필요하다

는 인식이 확산된 데 따른 것이다(황용수의, 2000).

ATP는 1988년 '일괄무역경쟁력법(Omnibus Trade and Competitiveness Act)'에 의해 개시되어 1990년부터 연방예산이 지원되고 있는데, 이 법에 명시된 프로그램의 목표는 (i) 중요 과학기술적 발견과 기술의 상업화를 촉진하고 (ii) 제조기술 고도화를 기하기 위하여 미국기업이 공유기술(Generic Technology) 또는 경쟁전단계 기술(Precompetitive Technology)의 개발과 산업적 응용을 도모하는 것을 지원하는데 있다(Ruegg, 1998). 1990년부터 지난 10년간 431개 연구팀이 프로그램에 참여하고 있다(<http://www.atp.nist.gov/99road/rslide14.htm>).

3. Framework 프로그램

Framework프로그램은 유럽연합(EU)의 대표적인 연구개발프로그램으로, EU가 지원하는 다양한 연구개발 프로그램들을 Framework 체제하에 편성하여, 여러 개의 구체적인 세부프로그램들로 구성되어 있다(<http://www.cordis.lu/f5>). 이 프로그램은 1984년에 시작되어 5차 프로그램이 1998-2002년까지 진행 중에 있으며, 유럽의 경쟁력을 강화하고, 회원국들간의 공동연구의 촉진이 주요 목표중의 하나이며, 복합기술을 대상으로 하고 있다.

기존의 Framework프로그램은 광범위한 과학기술분야를 종합적으로 지원하여 왔으나 현행 제5차 프로그램에서는 이러한 지원방식을 수정하여 중요 프로그램 분야에 집중하여 테마 프로그램(Thematic programmes) 및 구조 프로그램(Structural programmes)으로 발전시켜 나가고 있다. 사회적 측면(고용), 경제발전(경쟁력 강화), 유럽적 가치창출(개별국과의 공동협력을 통한 시너지 효과 창출) 등 세 가지 기준에 의하여 프로그램을 선정하고 소수의 핵심분야

2) ATP프로그램의 구체적인 내용들은 프로그램 홈페이지(<http://www.atp.nist.gov>)에서 공개하고 있다.

에 대한 선택과 지원을 강화하고 있다.

IV. 평가시스템 비교

본 장에서는 앞 장에서 논의한 세 프로그램의 평가시스템을 평가상황(Context), 집행 메카니즘(Mechanism), 평가결과(Outcomes) 및 결과활용(Utilisation) 등의 네가지 구성요소로 나누어 비교 분석하고자 한다.

1. 평가상황

세 프로그램 모두 평가가 중요하게 다루어진 점에서는 공통적이지만 각기 다른 배경을 가지고 있다. Alvey프로그램 평가는 초대 프로그램 책임자였던 Oakley의 결정으로 시작되었으며, 초기에는 그렇게 많은 기대를 하지 않았지만 의외의 성공³⁾을 거둔 사례의 하나이다(Oakley & Owens, 1989). 실제로 이 프로그램 평가는 프로그램을 성공적으로 이끈 견인차 역할을 한 것으로 알려져 있다⁴⁾. 반면 ATP평가는 외부의 특히 정치적인 많은 관심과 정치권의 프로그램 결과에 대한 요구에 대응하기 위하여 철저한 평가프로그램을 개발하고 있으며(Ruegg, 1998), 적극적이고 다양한 평가방법과 제도를 사용하고 있다는 점이 특징이다. 유럽의 연구개발프로그램 평가는 초기에는 단순연구에 대한 평가로부터 출발하였는데, 이후 Framework 프로그램의 지속적인 추진과 더불어 지식의 확산, 기술이전, 연구주체간 연계, 제품개발 및 상업화, 기타 기술혁신 지원서비스 및 하부구조 구축 등과 같은 활동을 포함하는 복잡한 연구활동에 대한

평가로 발전하였다(황용수의, 2000). 이러한 Framework 프로그램 평가의 초기 발전은 EU가 1994년 프로그램 평가를 보다 체계화하는 새로운 평가구조를 채택하면서 커다란 전기를 맞이하였다.

주요 평가관련자는 평가관리기관이나 요구기관 등인데, Alvey평가는 프로그램 사무국(Directorate)과 DTI(통산성) 등이고, 초기부터 그 운영과 결과에 대한 평가의 중요성을 깊이 인식하여 프로그램 수행기간 동안 평가작업이 계속 이루어지도록 하였으며, 평가결과의 피드백을 통해 프로그램 수행을 돕도록 하였다. 이러한 평가목적에 따라 실시간 평가(Real Time Evaluation)를 수행하였다는 점이 특징이다. ATP평가의 경우 평가관련자는 평가의 요구기관에 따라 외부와 내부요구로 구분할 수 있는데, 외부요구는 주로 상하의원, 의회 소위원회, GAO, OMB 등이며(Ruegg, 1996), 대부분이 정책적인 사안에 대해 제기되었고, 내부적인 요구는 상무성(DOC), NIST, 혹은 ATP자체에서 주로 프로그램 관리와 성과평가의 필요성에 의해 제기되는 등 다양하다. Framework프로그램의 경우는 평가가 제도화되어 있어, 프로그램 평가 및 세부프로그램 그리고 모니터링과 년단위 평가 등 복합적인 평가들이 체계적으로 동시에 진행되는 등 다양하다. 주요 평가관련자는 프로그램을 관리, 집행하는 유럽연합집행부(European Commission), 평가결과를 보고하고, 프로그램 관련 주요사항을 결정하는 유럽연합의회(European Parliament)와 각국 대표들로 구성된 European Council 등이다.

이처럼 평가관련자들의 평가에 대한 요구에 따라 평가의 기준이나 내용이 달라질 수 있다. 특히 평가관련자들에 가장 민감한 반응을 보인 프로그램은 역시 ATP라고 할 수 있다. Alvey의 경우 관련자들의

3) "의외의 성공"에 대하여는 Oakley의 회고록을(Oakley & Owens, 1989)바탕으로 한 연구자의 판단이다. 하지만 평가에 참여한 평가팀들의(Manchester 대학의 PREST, Sussex대학의 SPRU) 의견은 철저한 준비와 계획을 바탕으로 한 것으로 다른 평가를 하고 있다.

4) 프로그램에 대한 평가자의 적극적인 참여에 대하여는 프로그램 성과에의 기여라는 관점에서의 긍정적인 견해와 평가자의 포획(capture) 혹은 공정성의 의문 등의 관점에서의 부정적인 견해로 갈린다. 여기에는 관한 논의는 Ciba Foundation이 주최한 한 회의에서 이루어진 토론에 나타나 있다(Georghiou, 1989).

주된 관심은 기술개발이었던 반면, ATP는 프로그램을 통한 경제적인 파급효과에, 그리고 Framework의 경우는 유럽연합의 회원국간의 협력과 자원배분 등에 있어 평가의 초점도 여기에 맞추어져 있었다.

각 프로그램의 평가자는 크게 내부평가자와 외부평가자로 구분할 수 있는데, 평가자의 선정과정도 프로그램 평가의 특성을 반영하고 있었다. Alvey프로그램의 경우 평가자의 선정은 평가연구계획서를 바탕으로 한 공개경쟁의 과정을 거쳤으며, 최종적으로 3개의 대학연구팀을 평가자로 선정하고, 이들의 분야별 전문성과 팀간의 경쟁과 협력을 바탕으로 평가를 수행한 것이 특징적인 전략이라고 할 수 있다. Alvey 프로그램 평가는 처음부터 실시간 평가를 수행하였기 때문에 프로그램 수행 중에 프로그램의 관리와 운영을 개선하고자 하는 형성적 평가와 프로그램의 영향과 효율성을 평가하는 효용성 평가 모두에 초점을 맞추고 있었다. ATP의 경우 평가자와 평가의뢰 형태는 Weiss(1998)가 분류한 4가지 모든 유형이 다 사용되어, 그만큼 다양한 평가자들이 평가에 참여하고 있다는 것을 의미한다. 첫째 ATP는 내부 프로그

램 평가팀인 EAO(Economic Assessment Office)를 가지고 있으며, 여기서 프로그램 평가를 계획, 개발 및 집행하고 있다. 둘째로 여러 형태의 외부평가자와 컨설팅기관이 프로그램평가에 참여하고 있다. 셋째로 RFP(Request for Proposal)나 혹은 대학연구팀에게 평가가 의뢰되기도 하였다. 현재까지 미국경제학자들의 단체인 NBER(National Bureau of Economic Research)과 Jaffe(1996), Link(1996) 등 여러 학자들로부터 이론적인 지원을 받고 있다. Framework프로그램의 평가는 1996년부터 적용되어오고 있는 새로운 평가제도로 세부프로그램과 전체 프로그램을 위한 모니터링과 5개년 평가의 두 가지로 나뉘지고, 모두 독립적인 외부전문가에 의해 이루어진다. 모니터링에서의 외부전문가(패널)의 지원은 프로그램의 관리에 모니터링결과와 빠른 피드백을 위한 것으로 매년 주요 이슈에 대한 외부전문가의 의견을 듣기 위한 것이라고 한다면, 5개년 평가는 보다 광범위한 목표, 성과 및 실행 등을 평가하기 위한 것이다. 이들 외부평가자들은 그들의 개인적인 경험과 재능을 바탕으로 산업계, 학계, 연구기관 등의 균형 그리고 국가별, 혹은

<그림 2> 평가자의 위치

관 계	평가자유형	프로그램
내 부	피평가조직	Framework(Evaluation Unit)
	부처내 평가부서	
외 부	중앙관리부서	ATP(EAO)
	외부컨설팅기관	Framework(BETA, VanDijk B)
독립기관	대학 혹은 연구기관	ATP(CONSAD 등)
	커뮤니티와 소비자그룹	Alvey, ATP, Framework
	국립감사기관	ATP(NBER)
		Alvey(NAO), ATP(GAO)

5) 연구팀의 평가대상과 팀의 특성에 따라 혹은 이사회에 요구에 따라 다음과 같은 역할분담이 이루어졌다(Georghiou, 1991).
 ① PREST(The Policy Research in Engineering, Science and Technology, University of Manchester): 프로그램의 구조와 조직에 대한 평가. ② SPRU(The Science Policy Research Unit, University of Sussex): 프로그램의 효과성 평가, ③ CBS(The Centre for Business Strategy, London Business School): 영국 정보산업 각 분야에 미치는 영향 평가, 연구종료 후 1986년에 중단 탈퇴.

남녀의 성별 안배 등 균형있는 구성을 원칙으로 하고 있다. 평가자 선정은 EC(European Commission)에서 책임자에게 의뢰하는 형식으로 이루어진다. 한편 Alvey와 ATP평가의 경우 특징적인 사실은 감사기관인 NAO와 GAO가 각각 평가에 참여하였다는 사실이다. NAO는 Alvey프로그램에 대하여 2회, GAO는 ATP프로그램에 대하여 지금까지 4회 프로그램 평가를 수행하였으며, 프로그램에 대한 비판적인 평가보고서를 발행한 적이 있다. 지금까지 살펴본 평가자의 위치와 관련하여 <그림 2>와 같이 종합할 수 있다.

프로그램의 평가비용으로 Alvey는 실제값으로 전체 프로그램 예산의 0.28%가 평가예산으로 투자되었고(Guy et al, 1991), ATP는 5%, 그리고 Framework는 평가책임자의 추정에 의해 0.1% 정도가 평가를 위해 투자되고 있는 것으로 나타났다. 평가비용은 프로그램의 환경이나 평가의 범위에 따라 다양하다. 전체 프로그램 예산의 어느 정도가 프로그램 평가를 위한 비용으로 적합할 지에 대한 질문에 대해 25명의 면담자 중 22명이 1%정도가 적당하다는 응답을 하였다.

각 프로그램의 배경 및 특성에 따라 평가상황 역시 다르다. Alvey프로그램의 경우 프로그램 관리책임자의 요구에 의해 평가가 시작되고 복수의 외부평가팀을 구성하여 선의의 경쟁을 유도하고, 각 팀의 전문성을 살려주었으며, 특히 실시간 평가(Real Time Evaluation)라는 새로운 개념을 프로그램평가에 도입하게 된 계기를 마련하였다는 점에서 특징적이고, ATP의 평가는 프로그램의 정치적인 환경과 프로그램의 성과에 대한 많은 관심으로 자연히 평가에 적극적이었고, 다양한 평가자 및 평가방법이 동원되었다는 점에서 주목할 만하다. Framework 프로그램 평가는 다국간 공동연구프로그램의 복잡성으로 인하여 평가를 위한 전문적인 관리체제를 갖추지 않을 수 없었다. 특히 유럽에의 풍부한 평가전문가들을 활용

한 철저한 평가시스템을 구축하고 있고, 지속적이고 체계적인 평가가 이루어질 수 있도록 정교한 제도적인 장치를 마련해두고 있는 점이 특징이다.

Alvey프로그램의 경우는 실시간 평가라는 프로그램에 실시간 관찰자로서 평가자가 참여하는 가장 적극적인 방법을 사용하고 있고, Framework프로그램 역시 모니터링과 평가를 혼합하여 지속적이고 체계적인 평가(Continuous and Systematic Evaluation)가 이루어지는 평가시스템을 가지고 있다. 한편 Alvey와 Framework 프로그램이 외부평가를 하고 있는 반면, ATP프로그램은 내부평가와 외부평가를 혼합하고 있고, 평가제도의 우수성보다는 다양하고 복합적인 평가제도의 시험장으로서, 그리고 평가에 엄청난 투자를 하고 있다는 점에서 주목을 받고 있다(Georghiou & Roessner, 2000).

2. 집행메카니즘

평가의 집행메카니즘에 관하여는 앞장에서 살펴본 바와 같이 평가기준, 평가방법 및 의사소통 메카니즘에 초점을 두고 각 프로그램의 특성들을 비교 분석한다. Alvey프로그램의 평가기준들은 프로그램의 집행, 진척도, 그리고 적절성 등으로 이루어져 있고(Guy et al, 1991), 프로그램의 집행에 관하여는 프로그램 운영조직, 자금지원, 공동연구 현황, 모니터링 시스템, 다른 프로그램과의 관계 등의 항목으로 구성되어 있다. 프로그램의 진척도와 영향에 관하여는 기술진보, 상업화 가능성 또는 연구결과의 배포, 학문연구에의 영향, 기업전략에의 영향, 연구수행 및 연구인력에 미치는 영향 등을 평가하고 있다. 프로그램의 적절성에 관하여는 기술분야, 전략, 정부지원 수준, 국제적 상황, 경쟁전 연구의 적절성, 미래의 대체적인 경향의 관점에서 평가한 프로그램의 적절성 등에 대

6) 면담과정 중에 본 연구자의 질문에 대한 대답을 위해 평가책임자가 추정 계산한 것임.

해 평가가 이루어져 프로그램의 국가적 목적 및 관리에 많은 비중을 두고 있음을 알 수 있다. ATP프로그램의 경우는 경제적인 영향평가에 많은 비중을 두고 있다. 기술적용의 경제적 가치, 기술의 비용절감효과 등 미시경제적 측면과 프로그램의 국민경제에 미치는 영향 등의 거시 경제적 측면에서의 평가 및 ATP의 자금지원에 의한 벤처기업과 비지원 벤처기업간의 특성 비교 등이 이루어진 바 있다(Ruegg, 1998). 앞서 논의한 바와 같이 ATP평가에는 다양한 외부경제학자들의 참여가 있는데, Jaff(1998)의 연구개발의 사회환원율(Social Rate of Return) 측정을 위한 개념을 적립하였고, 연구개발의 효과를 지식파급효과(Knowledge Spillover), 시장파급효과(Market Spillover), 그리고 네트워크 파급효과(Network Spillover)의 세 가지로 구분하고 측정 및 비교하였다. 한편 Link(1998)는 프로그램지원으로 인한 비용감소에 대한 평가를 한 것 등이 특징적이다. 유럽연합(EU)은 연구성과와 혁신에 기여, 유럽공동체의 가치 창출, 사회적 목표에 기여, 경제발전과 과학기술의 미래 비전(Vision) 제시, 자원의 효율적인 활용, 연구주체간 파트너쉽 강화, 프로그램 관리의 개선 등 프로그램의 평가기준을 제시한 기본 지침을 제정하였으며, 연구개발프로그램에 대한 EU의 평가기준은 유럽의 산업경쟁력 강화를 도모하는 것을 중시하였으나, 1990년대 후반부터는 고용, 환경, 의료 등의 사회적인 목표에 대한 기여를 강조하고 있다. 이처럼 평가기준 역시 프로그램의 배경이나 특성을 반영하고 있고 이에 따라 다양한 기준들이 사용되고 있다.

프로그램 평가를 위해서는 여러 가지 평가방법을 혼합하여 쓰는 것이 보통인데 프로그램의 특성에 따라 적절한 평가방법의 혼합 역시 달라진다. Alvey프로그램은 다양한 방법들이 사용되었는데, 대표적으로는 설문조사 방법, 구조화된 면접, 특히 서지분석(Bibliometric Analysis), 사례연구, 역사적 연구, 국제비교 등이 주로 이용되었다(Hobday, 1988). 특히 설문조

사시 프로그램에 대한 비우호적인 집단을 포함시킨 것도 또 하나의 방법상의 특징이라고 할 수 있다. Alvey프로그램 평가의 정보수집상의 특징은 실시간 평가를 통해 프로그램 종료 후에는 얻을 수 없는 정확하고 실제적인 자료와 전문가 및 다른 참가자의 현장 의견을 수집할 수 있고, 적절한 자료를 적절한 시점에 확보할 수 있었다는 점이다. 그러나 실시간 평가는 과제 선정에서부터 중간점검을 거쳐 종료 후의 영향까지를 평가의 대상으로 하고 있어 사후평가나 사전 평가보다 많은 비용이 든다는 점과 프로그램 추진과정에 평가팀이 참여함으로 인해 평가의 공정성을 해칠 우려 등 단점이 지적되었다(Georghiou, 1989).

ATP평가의 방법론상의 중요한 특징 중의 하나는 체계적인 자료 수집 분석을 위해 복합적인 접근방법을 사용한다는 데 있다. ATP와 같은 종합적인 프로그램의 경우, 프로그램의 내부적인 관리상의 요구와 의회, 산업계 등의 외부로부터 제기되는 다양한 요구에 대응하기 위해서는 많은 평가방법의 도움을 받지 않으면 안된다(황용수외, 2000). 이와 함께, 실증적인 장기적 프로그램 산물을 평가하는데 오랜 시간이 걸린다는 점을 감안하여 평가결과에 대한 단기적인 긴급요구도 반영할 수 있는 평가방법도 강구되어야 할 필요가 있다는 것이다(Ruegg, 1996). ATP는 처음부터 신청자, 프로젝트, 실현성과 및 당면과제 등을 지속적으로 기록하여 일상적인 관리와 통제 및 자문을 하고, 외부평가자에게 제공할 평가정보를 지속적으로 수집하여 활용할 수 있는 정보수집 및 기반이 확보되어 있다. Framework프로그램 평가에서의 정보수집은 모니터링을 통하여 지속적으로 평가를 위한 실시간 정보를 수집하고, 수집된 정보가 Framework 프로그램의 연차별 모니터링과 5개년 평가에 활용되도록 하고 있다. 모니터링과 5개년 평가에 적용되는 평가기준은 상당한 차이가 있다. 모니터링에서는 (i) 프로그램 운영의 효율성, (ii) 선정된 프로젝트들이 EU의 정책목표와 어느 정도 일치하는지 여부, (iii)

구체적인 측정방법 사용, (iv) 설정된 목표의 달성 정도, 그리고 (v) 변화하는 환경의 관점에서 사회의 요구에 대한 신축적 대응 여부 등이 중요한 평가기준이다. 5개년 평가는 프로그램의 전략, 집행, 성과 또는 영향에 대한 전반적인 내용을 다루게 되고, 이에 따라 평가기준도 다음에서 제시하는 (i) 적합성, (ii) 효율성, (iii) 효과성 등과 같이 보다 포괄적인 것이다. 한편, 프로그램 차원에서의 평가에서는 EU의 과학기술정책과 개별 국가의 정책 간의 연계성 여부를 중요한 평가기준으로 삼고 있다. 현행 제5차 Framework 프로그램에서는 사회경제적인 목표를 강조함으로써 보건, 환경, 고용, 근로여건 등과 관련된 사회경제적인 영향을 평가하여 제시할 수 있는 새로운 방법론의 개발이 요즘 유럽연합의 평가전문가들에게 관심의 대상이다.

평가과정에서 평가자와 평가관련자들간에 다양한 목적으로 의사소통이 이루어진다. Alvey프로그램은 평가자와 프로그램관리조직이나 참여연구자들과의 다양한 의사소통채널을 가지고 있었다. 일년에 두 번 운영위원회에서 평가결과와 향후방향에 관해 토의하였고, 분기별로 실무위원회에서 평가의 관리상의 문제들을 협의하였다. 통산성(DTI)의 경제부는 평가보고서나 평가방법상의 문제들을 논의하기 위하여 분기별로 워크숍을 개최하였다. 평가보고서들은 발간되기 전에 관련자들의 회람을 통하여 의견을 수렴하였다. 그러나 통산성(DTI), 국방성(MoD), SERC 등 정부부처간의 상호교류와 커뮤니케이션은 잘 이루어지지 않았던 점이 프로그램 평가에서 자주 지적되었던 부분이다. 특히 프로그램평가과정에서 특징적인 것은 평가팀은 매년 회의개최와 공식모임이 끝난 후 연구자들과의 비공식적인 모임을 통한 정보수집 및 의견교환이 평가자들의 프로그램진행상의 상황판단과 연

구자 및 관리자들의 의견파악에 실질적인 도움이 되었다는 것이다(인터뷰, 1999. 12).

ATP프로그램 평가의 경우도 내부평가자와 연구자간의 실시간 모니터링에 의한 대면(Face to Face)의 의사소통 혹은 진도보고시스템(BRS, Business Reporting System)에 의해서 내부평가자간의 상호정보의 교류가 이루어지고 있으며, 다양한 평가지원시스템에 의해 외부평가전문가나 경제전문가들과의 커뮤니케이션 메카니즘이 형성되어 있다. 그러나 여러 가지 형태의 의사소통 메카니즘이 존재함에도 불구하고 Alvey프로그램이나 Framework프로그램과 같은 자연스러운 평가결과의 피드백 메카니즘은 찾아볼 수가 없다는 것이 아쉽다.

Framework프로그램의 많은 회원국들이 참여하고 있는 복잡한 프로그램의 구조와 광범위한 대상연구영역 등으로 그 의사소통 메카니즘 역시 복잡하다. 평가패널과 의사결정자간 혹은 패널간의 의사소통은 계속적이고 체계적이다. 전체적인 의사소통은 연구개발DG의 평가국(Evaluation Unit)이 패널모임 및 인터뷰 등을 조정하고 운영하며, 평가를 위해 필요한 데이터와 정보를 제공하는 등의 지원기능을 한다(European Commission, 1999).

Alvey프로그램은 평가자와 정책결정자간의 실시간 밀접한 의사소통을 통해 평가결과의 피드백이 순조롭게 이루어진 반면, 프로그램관련 정보들에 대한 DB 구축이 늦어져 평가자들로부터 비판을 받았다. 프로그램평가를 위한 데이터베이스의 구축이 늦어 실질적인 실시간 평가에 많은 장애와 한계가 있었으며, 이것은 프로그램이 끝난 이후까지도 지적되었던 문제중의 하나이다. 반면 ATP는 초기부터 내부평가자의 실시간 모니터링이나 BRS 등 제도적인 장치를 통해 프로그램에 관한 정보를 수집하고 데이터베이

7) 일반적인 프로그램 관련 자료수집방법으로는 BRS(Business Reporting System)에 의한 통계자료 수집, 제3의 기관을 활용한 조사 및 ATP의 특

정 사례연구 등을 이용하고 있다.

스를 구축하는데 적극적이었다. 그러나 다양한 평가자와 평가시스템에도 불구하고 결과의 피드백을 위한 커뮤니케이션 메카니즘이 부족하다는 한계를 지니고 있다.

3. 평가결과

Alvey프로그램의 경우 프로그램의 시작부터 실시간 평가를 하여 평가의 결과를 수시로 프로그램 관리자나 스텝에게 피드백하여 프로그램의 성과와 지휘에 많은 영향을 미쳤다. 실제로 30여종 이상의 실시간 평가보고서가 수시로 혹은 연구진행계획에 따라 이사회에 제출되었으며, 이들을 종합한 중간보고서와 최종보고서가 발간되었다. 구체적으로 프로그램과 관련된 주제별 보고서, 1987년까지의 중간보고서(Guy et al, 1987), 그리고 프로그램 종료 1년 후인 1991년의 최종평가보고서(Guy et al, 1991) 발간이외에 정기적인 공식 비공식적인 프로그램사무국과의 회의를 통해 평가결과를 보고하고 정책제안 등이 이루어졌다.

평가결과와 관련하여 ATP평가의 특징적인 사실은 다량의 다양한 평가보고서의 발간이며, 이에 따른 긍정과 부정적인 시각이 동시에 있다. 다양한 평가의 요구와 복합기술프로그램의 특성상 복합적인 평가시스템을 갖추고 있다는 것이 긍정적이라면, 평가에 대한 과도한 투자와 노력, 그리고 너무 일찍 프로그램의 영향측정을 시도하여 프로그램이 정착되기도 전에 얼마나 자랐는지 뿌리를 파보는 경이라는 비판도 있다(Goldston, 1999). 특히 ATP평가는 성과의 합산이나 결과측면의 평가보다는 관리자들이 직접적인 문제를 해결하는데 도움이 될 수 있도록, 현재 프로그램 운영에 대한 과정적 정보를 제공하는데 초점을 두어 유용한 정보가 빠르게 발생하도록 하였다는 점에서 좋은 인식을 심어주고 있다(Ruegg, 1998).

Framework프로그램 평가의 결과보고서는 앞서 논

의한 바와 같이 매년 연차별 모니터링 보고서와 5개년 평가보고서가 각각 세부프로그램들과 전체프로그램에 대하여 발간된다. 특히 모니터링 보고서에는 평가대상기관이나 프로그램관리자의 보고서에 대한 의견을 동시에 신도록 함으로써 평가과정에서의 신중성과 프로그램 관리자의 의사결정에 도움을 주도록 하고 있다. 각 평가보고서는 EC 및 유럽의회에 보고되어 정책결정에 중요한 자료로 활용된다. 앞에서 본 바와 같이 모니터링 및 5개년 평가의 기본목적은 프로그램 관리를 개선하고 집행의 효율성을 향상시키며 차기 프로그램을 기획하는데 기초자료로 활용하기 위한 데 있다.

평가결과의 보고서와 관련하여 주목할 만한 것은 미국 GAO식 평가보고서 발간이다. GAO평가보고서에는 보고서가 발간되기 전에 관련기관의 의견을 물어 수정하거나 의견이 일치하지 않을 경우 보고서에 포함시키는 등의 사전 '피드백 과정'을 거치도록 하고 있다. ATP평가의 경우 역시 그들의 평가결과에 대한 상무성(DoC)의 의견과, 또 상무성의 의견에 대한 GAO의 견해를 동시에 담도록 함으로써 평가에 대한 신중성과 공정성을 기하고 있다. 현행 Framework프로그램의 모니터링보고서도 보고서에 대한 프로그램책임자의 견해를 신도록 하고 있다.

평가결과보고와 관련하여 특징적인 사실은 Alvey 프로그램의 경우는 실시간평가를 하고 있어 정기적인 평가보고서보다는 당시의 주요 이슈별로 비정기적인 조사 및 결과보고를 위한 회의 등이 더 빈번히 이루어졌다는 것이 차별되는 특징이다. ATP프로그램 역시 다양한 평가결과가 보고되지만 Alvey프로그램에 비해 평가보고서의 발간 외에 결과보고를 위한 행사가 특별히 이루어지고 있지는 않다. Framework 프로그램 역시 체계적이고 지속적인 평가가 이루어지도록 제도화하여 1년 혹은 5년 단위의 보고서와 그 평가결과가 상호연계되도록 하고 있다는 점이 Alvey와 ATP의 두 프로그램에 비해 특징적이다.

4. 결과활용

프로그램평가 결과의 활용을 위한 전략중의 하나는 평가자와 프로그램관리자 혹은 평가자와 연구자 사이의 결과활용을 위한 활발한 의사소통 메카니즘을 구축하는 것이다. Alvey프로그램의 경우는 실시간 평가로 특별히 결과활용을 위한 의사소통 구조를 구분하기 어렵지만 평가자들과 프로그램관리자 간에 수시로 결과보고 및 활용을 위한 의견교환의 장이 마련되어 있었으며, 1년에 두 번씩 개최되었던 프로그램 책임자를 포함한 운영 위원회 (Steering Committee)에서 평가의 결과와 향후 진행방향에 관한 논의가 많이 이루어졌다. 특히 프로그램의 주관기관인 통산성(DTI) 공무원들이 직접 평가팀에 참여하여 결과의 정책적 활용이 자연스럽게 이루어질 수 있는 구조를 가지고 있었다.

ATP의 프로그램 평가의 원칙은 평가를 프로그램의 동적인 구조와 통합하여 지속적인 프로그램관리 개선과 연계시킨다는 전략이다. 따라서 중간평가 혹은 실시간 평가를 통해 얻어진 정보들이 프로그램에 피드백될 수 있도록 하고 있다. 다만 평가의 많은 부분을 프로그램의 생존(Survival)에 비중을 두고 있었기 때문에 프로그램의 관리 혹은 정책제안 등의 실질적인 결과활용을 위한 의사소통 메카니즘은 찾아보기 어렵다.

Framework프로그램의 경우 평가보고서를 통해 전문가들이 제시한 정책제안에 대하여는 EU가 철저한 검토과정을 거쳐 내부적으로 그 반영 여부를 결정한다. 그리고, 평가결과는 차년도의 모니터링이나 평가과정에서 검토되도록 하여 평가과정이 평가보고서의 발간으로 그치는 것이 아니라 정책적인 피드백을 가져오는 제도적인 메카니즘을 만들어 두고 있다. 그런데 실제 평가결과가 집행상의 문제점을 개선하는 효과를 달성하였는지는 별도의 검토가 필요하다고 할

수 있다. Framework프로그램의 경우 평가결과의 활용도를 높이기 위하여 EU위원회로 하여금 평가 및 모니터링 결과와 그들의 정책제안에 대하여 수용여부를 결정하도록 하고 있다. 한편 EC의 연구성 (Research DG)의 평가국(Research Evaluation Unit)이 주관하는 평가회의를 매년 당시 유럽연합의 의장국에서 개최하여 의장이 참여하도록 하고, 그 의견이 반영되도록 하는 등 적극적이다.

평가결과의 대외공개에 관하여는 세 프로그램 모두 평가보고서의 공개를 원칙으로 하고 있으나 가장 적극적으로 프로그램 평가결과를 대외에 공개하는 것은 ATP프로그램이다. ATP의 진행평가에 관한 모든 보고서는 인터넷이나 문서를 통하여 일반에 공개하고 있고, 결과는 NIST스텝, 프로그램관리자 및 미래의 과제제안자에게 유용한 정보원이 되고 있으며, 기술혁신을 위한 정책 및 프로그램 개발에 대하여 이들이 점진적이고 실증적인 태도를 취할 수 있도록 도와 주고 있다. Alvey프로그램의 경우는 중간보고서와 최종보고서가 공개되어 있으나, 그의 많은 실시간으로 보고된 평가보고서들은 평가자와 프로그램책임자 혹은 통산성(DTI)간의 직접적인 보고형식을 취해 공개되지 않은 것들이 대부분이다. Framework프로그램 역시 평가결과의 공개에는 적극적이다. 년차별 혹은 5개년 각각의 모니터링보고서와 평가보고서가 원칙적으로 공개되고 있으나 ATP의 경우처럼 인터넷상에 보고서 전체원문제공이 이루어지는 경우는 많지 않다.

평가결과의 직접적인 영향 외에도 간접적으로 후속평가에 영향을 미치고 있는데, Alvey 프로그램평가에서 사용된 새로운 평가기법들은 이후 영국뿐만 아니라 유럽연합을 비롯한 다른 나라의 프로그램평가에 영향을 미치고 있다(Lukkonen, 1999). Alvey 프로그램평가의 가장 두드러진 학문적인 그리고 실질적인 기여는 실시간 평가의 개념정립 및 그 적용이라고 할 수 있다. 평가의 결과에 대하여는 경쟁적 상충적인 시각을 가지는 경우가 많고, 프로그램 평가의

경우 결과의 활용에 관하여 소홀히 다루어지는 경우가 많다. 따라서 결과의 활용은 매우 정치적이고, 미국의 중소기업혁신연구 프로그램과 같이 이해관계자들의 로비능력에 의해 결과가 공개되지 않는 등 프로그램의 상황에 따라 다양하다(Eveland, 1986).

V. 정책적 시사점 및 한계

지금까지 살펴 본 각 프로그램 평가들의 정책적 시사점들을 살펴보면 Alvey 프로그램 평가는 현재에도 영국을 비롯한 EU 프로그램 평가에 많은 영향을 미치고 있다.

하지만 Alvey 프로그램 평가상의 비판은 우선 대부분 프로그램이 갖는 부처간의 조정의 문제가 항상 제기되었고, 실제로 국방성의 참여는 미온적이었으며, 정치권의 초기지원을 받지 못한 점등이 문제로 지적되었다. 평가와 관련된 직접적인 문제는 외부 독립 평가자들의 프로그램 운영 및 관리에의 지속적인 참여에 따른 평가의 공정성 시비였다. 특히 이들 평가자들의 프로그램에의 영향이 커지면서 평가자들에 대한 곱지 않은 시각이 있었고, 이런 배경에는 평가자들이 후속 프로그램의 설계에의 적극적인 참여와 특히 기자회견 등을 통한 Alvey 프로그램 지원 등으로 외부 독립평가자들의 내부화에 대한 우려가 있었다. 한편 앞에서 지적한 바와 같이 실시간 평가로 인한 다수 평가자들의 동원 및 활동으로 과잉투자의 논란이 있었다.

한편 ATP의 프로그램 평가가 주목받는 이유는 그 평가시스템의 우수성이나 평가방법상의 특징 때문이라기보다는 다양하고 적극적인 평가시스템의 구축과 그 평가를 위한 엄청난 노력과 투자때문이다(Georghiou & Rossner, 2000). 따라서 지금까지 많은 다양한 평가보고서가 발행되었으며, 이러한 다양한 평가시스템의 관점에서 ATP 평가가 갖은 특징중의 하나는 RFP

혹은 대학연구팀의 평가에의 참여유도를 통한 프로그램 평가의 지원이다. 경제학자들의 참여와 특히 미국경제학자들의 단체인 NBER과의 지속적인 이론적인 지원 등이 특징이다.

ATP 평가에 있어서의 연구자와 내부평가자와의 실시간 모니터링을 통한 커뮤니케이션 역시 구별되는 특징중의 하나이다. 모든 프로젝트는 ATP 내부평가자에 의해 실시간 모니터링(Real time monitoring)되고 있으며, 이를 통해 프로그램에 참여하고 있는 연구자들과 평가자들 간의 자유로운 직접적인 의견과 정보의 교환이 이루어진다는 점이다.

반면 ATP 평가의 한계는 프로그램의 정치적인 생존을 위한 평가에 많은 비중을 두고 있어 프로그램에 대한 계량적인 성과 혹은 경제적인 효과측정에 초기부터 과잉측정을 한다는 비판이 있어왔다. 특히 의회의 요구에 의한 프로그램을 평가한 것이지만 GAO의 ATP의 평가와는 견해차이가 많고, 실제로 근본적인 프로그램의 성과에 대한 것에 있어서는 프로그램 운영기관인 NIST 혹은 주관정부기관인 DoC(상무성)과 GAO 평가간에 상당한 의견차이가 있어 평가의 공정성에 의문이 제기되기도 했었다.

유럽연합의 대표적인 연구개발프로그램인 Framework 프로그램 평가의 특징은 그 평가시스템이 이론적 실제적으로 철저하다는 것이다. 이러한 배경에는 우선은 국제적인 프로그램으로 많은 관련 EU 국가들의 관심이 쏠려 있기 때문에 거기에 부응해야 하고, 두 번째는 유럽전역의 풍부한 평가전문가들이 프로그램 평가에 참여하고, 또 시스템을 개발하고 있기 때문에 체계적인 평가제도를 갖출 수 있었던 것 같다. 초기에는 연구성(Research DG)에서 주로 프로그램을 평가하고, 개발해왔으나, 최근들어 모든 부처(DG)에 평가담당부서를 만들고, 예산성(Budget DG)에서 종합하도록 하고 있으며, 각 부처의 평가담당자들간의 네트워크를 형성하여 그 평가결과가 예산에 반영될 수 있도록 하고 있다. 감사성(Audit DG)의

경우는 이들 평가에 대한 메타평가를 하도록 함으로써 구조적으로 철저한 평가시스템을 갖추기 위한 많은 노력이 이루어지고 있다는 것을 알 수 있다.

Framework프로그램 평가의 한계는 국제적인 프로그램으로서 갖는 일반적인 문제와 특성들을 내포하고 있다. 특히 연구비, 참여연구자 등 모든 측면에서 국가별 안배가 중요하게 고려하고 있고, 실질적인 연구성과에 앞서 국가간 공동연구를 강조하고 있다는 것 등이 단일국가의 전략적인 프로그램과 그 평가구조와는 차이가 있다.

프로그램 평가는 사회적 과정이며, 국가 사회적 환경의 영향을 받게 된다. 특히 국가의 전략연구개발프로그램의 경우 대부분 정치 및 언론에 많은 관심을 받아 평가외적인 영향을 받게 되고, 이로 인한 프로그램 평가의 특성이 달라지기도 한다. 앞서 살펴본 바와 같이 이들 프로그램들은 그 형성배경 및 국가 사회적 환경에 따라 다른 평가 제도와 전략을 사용하고 있다. 그러나 우리에게 주는 시사점으로 일반적인 특징은 세 가지 사례 모두 프로그램의 형성과 동시에 평가에 대한 계획이 이루어지고, 프로그램의 진행과 프로그램의 평가를 병행함으로써 프로그램 관리 및 정책에 반영이 이루어지도록 제도적 장치뿐만 아니라 실질적으로 철저한 시행을 하고 있다는 점이다. 이러한 특징을 나타내는 한마디는 곧 '지속적이고 체계적인 평가(Continuous and Systematic Evaluation)'로 표현된다. 이것은 곧 사후결과에 대한 종합 평가보다는 프로그램의 과정평가에 그리고 그 평가결과의 피드백에 평가의 많은 비중을 두고 있다는 의미이다. 우리 나라의 경우 우선 종합적인 프로그램 평가가 이루어지는 경우가 드물고, 평가가 프로그램의 마무리를 위한 한 과정으로서 이루어지는 경우가 많으며, 그 결과의 정책적 반영이 잘 이루어지지 않는 문제점들을 안고 있는 환경에서 특히 이들 프로그램 평가가 시사하는 바가 크다.

참고문헌

- 황용수의 (2000), 「정부연구개발프로그램체계의 비교 분석과 향후평가체계구축방안」, 과학기술부.
- ATP (1999), *General Terms and Conditions - Advanced Technology Programme*.
- Brinkerhoff, R.O., D.M. Brethower, T. Hluchyj & J.R. Nowakowski (1983), *Programme Evaluation: a Practitioner's Guide for Trainers and Educators*, Boston, Kluwer-Nijhoff.
- Dumont, Y., L. Durieux, L. Karatzas, L. O'Sullivan, H. Teuber, G. Stroud & G. Fayl (1999), *EU RTD Programmes Impacts/Results Assessment: a Manifold Task in a Complex System. Evaluation of Science and Technology in the New Europe*, Berline, European Commission.
- Frame, J. D. (1983), "Quantitative Indicators for Evaluation of Basic Reserach Program and Project", *IEEE Transactions on Engineering Management*, 30(3).
- GAO (1992), *The Evaluation Synthesis*.
- GAO (1996), *Measuring Performance: the Measuring Performance: the Advanced Technology Program and Private-Sector Funding*.
- Georghiou, L. (1989), *Orgaization of Evaluation. the Evaluation of Scientific Research-Ciba Foundation Conference*, John Wiley & Sons.
- Georghiou, L. (1987), *Evaluation of Research*, OECD, Paris.
- Georghiou, L. (1995), "Research evaluation in European National Science and Technology Systems", *Research Evaluation*, 5(1): 3-10.
- Georghiou, L. & D. Roessner (2000), "Evaluating Technology Programs: Tools and Methods",

- Research Policy*, 30: 891-903.
- Goldston, D. (1999). "Perspectives on the Programme." in Wessner, C.W.(ed), *The Advanced Technology Programme: Challenges and Opportunities*.
- Guy, K., P. Georghiou, L. Quintas, M. Hobday, H. Cameron & T. Ray (1991), *Evaluation of Alvey Programme for Information Technology*, London: HMSO.
- Guy, K., L. Georghiou, L. Quintas, M. Hobday, H. Cameron & T. Ray (1987), *Evaluation of the Alvey Programme-Interim Report*, London: HMSO.
- Hill, C. (1999), "An Overview of the ATP's History and Objectives", in Wessner, C.W.(ed), *The Advanced Technology Programme: Challenges and Opportunities*.
- Hobday, M. (1988), "Evaluation Collaborative R&D Programmes in Information Technology: the Case of the UK. Alvey Programme", *Technovation*, 8: 271-298.
- Jaffe, A.B. (1998), "The Importance of Spillovers in the Policy Mission of the Advanced Technology Program", *Journal of Technology Transfer*, 23(2): 11-19.
- Joint Committee on Standards for Educational Evaluation (1994), *The Programme Evaluation Standards, Thousands Oaks, Sage Publications*.
- Link, A.N. (1993), *Advanced Technology Program: Economic Study of the Printed Wiring Board Joint Venture After Two Years, ATP*.
- Luukkenen, T (1999), *Additionality of Publicly-Funded RTD Programmes*, The ITPS Report, 40.
- Oakley, Brian, Kenneth Owen (1989), *Alvey: Britain's Strategic Computing Initiative*, The MIT Press.
- Roessner, D. (1989), "Evaluating Government Innovation Programmes: Lessons from the US Experience", *Research Policy*, 18(6): 343-359.
- Ruegg, R. (1996), *Guidelines for Economic Evaluation of the Advanced Technology Program*, NIST.
- Ruegg, R. (1998), "The Advanced Technology Programme, Its Evaluation Plan, and Progress in Implementation", *The Journal Technology Transfer*, 23(2): 5-10.
- Scriven, M. (1991), *Evaluation Thesaurus*, 4th, Newbury Park, CA, Sage Publications.
- Suchman, E. A. (1972), "Action for What? A Critique of Evaluative Research", in Weiss, C. H.(ed), *Evaluating Action Programs*, Allyn & Bacon.
- Weiss, C.H. (1972), *Evaluation Research: Methods of Assessing Program Effectiveness*, Engle Wood Cliffs, Prentics-Hall Inc.
- Weiss, C.H. (ed) (1977), *Using Social Research in Public Policy Making*, Lexington Books.
- Weiss, C.H. (1998), *Evaluation-Methods for Studying Programs and Policies*, New Jersey, Prentice Hall.