

# 국책 연구개발 프로그램에 대한 외국평가사례 분석 및 평가지침 개발†

이무신\* · 엄기용\*\* · 심덕섭\*\*\*

Evaluations of National Innovation/R&D  
Programs in the Developed Countries and Its Implications†

Mushin Lee\* · Kiyong Om\*\* · Deogseob Shim\*\*\*

## ABSTRACT

Regarding the study in innovation/R&D programs evaluation, twelve cases from the U. S., U.K., Germany, France, Sweden, Australia, Japan, and EC are analyzed, and its implications to Korean situation are investigated. From the analysis of the cases, some valuable guidelines for Korea are developed. Some of the guidelines are: the evaluation must be treated as an integral part of the Government supported programs, specialists in evaluation should be raised, the long-term effects and non-economic/social effects should also be included in the evaluation, various efforts to acquire more exact and versatile information necessary for quality evaluation should be made, and additional effort is required for the evaluation results to be actualized.

## 1. 서 론

1970년대부터 OECD 회원국들을 중심으로 많은 국가들이 산업에서의 혁신 및 연구개발에 관심을 가져왔고, 이를 촉진하기 위한 구체적인 방

안으로 각종 법안과 프로그램을 기획·실행하였다. 미국에서는 지속되는 예산적자와 국제경쟁력 하락을 개선하고자, R&D 세금공제 프로그램을 비롯한 다양한 종류의 혁신 진흥 프로그램을 시작했으며, 유럽에서도 경기회복을 위해서는 혁신

† 본고는 1993년에 정보통신연구관리단이 지원하고 집필자 등이 수행한 "정보통신연구개발사업의 종합평가모형 개발에 관한 연구"의 일부분을 정리하게 제시한 것이다.

\* 한국과학기술원(KAIST) 경영정책학과 교수

\*\* 한국과학기술원(KAIST) 경영정책학과 박사과정

\*\*\* 한국과학기술원(KAIST) 경영정책학과 석사과정

과 연구개발이 필수적임을 인식하고 지속적으로 그 규모 및 종류를 확장해왔다[30]. 대표적으로 기업의 혁신활동에 대한 보조금 지급, 산업에서 필요한 특정기술개발 프로젝트 지원, 대학 및 연구소에서 개발된 기술의 이전을 촉진하기 위한 프로그램 개발, 기업의 혁신을 위한 정보제공 및 자문활동 촉진 등 다양한 정책을 개발하고 이에 막대한 자본과 인력을 투입하고 있다[23].

그러나 이러한 정책과 프로그램 활동의 이면에는 아무런 실증적 증거나 지식이 없었다. 첨단기술분야의 경쟁력 제고와 국가경제 활성화를 위해서는 기술혁신이 필수적이라고 믿어왔으며, 연구개발 프로그램은 다른 사회적 문제 해결을 위한 프로그램과는 다르다고 인식하여 공식적인 정부의 평가과정에서 제외시키기도 하였다[8]. 그러나 국가간 기술경쟁이 첨예화되고 연구개발 비용이 기하급수적으로 증가되는 반면, 이용가능한 예산과 자원에는 제약이 있으므로, 정부와 국민은 투입된 자원에 대하여 사용의 정당성 뿐만 아니라 그에 합당한 성과를 요구하게 되었다. 1980년대에 들어와서는, 혁신정책에 대한 신념만 있지 이해가 부족하다는 결론아래, 실증적인 자료를 수집하고자 노력하였다. 영국은 평가를 프로그램의 일부로 포함시켰으며, 프랑스는 11개 연구조직을 평가하여 연구잠재력과 경제적 현실간의 거리를 좁히려고 노력하였고, 오스트레일리아와 일본도 재정감축과 함께 평가노력을 증대하였다[30]. 또한 OECD 회원국들은 공식적 모임을 통하여 혁신 프로그램의 평가방법, 평가절차 그리고 기획과 실행 및 결과의 활용이 적절히 이루어진 평가상황 등에 대한 서로의 평가경험을 공유하였다[25].

이러한 상황은 우리나라에도 그대로 적용된다. 우리나라의 국책연구개발사업은, 국민복지증진에 직접적으로 기여할 수 있는 공익성이 큰 공공기술개발과 민간의 능력으로는 수행하기 어려운 핵

심기술분야의 기술개발에 선도적인 역할을 담당하여 왔다. 이러한 국책연구개발사업을 보다 효과적이고 효율적으로 추진하기 위해서는, 국가 연구개발자원의 투입을 확대하여야 할 뿐만 아니라, 투입된 자원에서 최대의 성과를 산출하도록 자원의 효율적 관리를 위한 노력을 반드시 병행하여야 한다. 자원의 효율적 관리를 위한 필수적인 관리기능 중의 하나가 연구개발 평가기능이다[1].

이처럼 세계 각국은, 국책연구개발사업의 확대, 수정 및 중단결정, 그리고 사업관리의 효과성과 자원 사용의 효율성 향상을 위하여, 국책연구개발사업의 평가에 지대한 관심과 노력을 기울이고 있는 것을 알 수 있다. 국가적 차원에서 연구개발 평가가 중요해지는 이유로는 다음 네가지를 들 수 있다[4]. 첫째, 국가간 기술개발 경쟁이 첨예화되어, 자원의 효율적 사용을 위한 연구영역, 연구과제 등의 우선순위를 명확히 하는 것이 필요해졌다. 둘째, 기술환경이 복잡하고 연구비가 대규모화됨에 따라 연구개발의 실패가능성과 충격이 거대화되어, 상황 변화에 신속하게 대응하는 관리기법에 대한 요구가 대두되었다. 셋째, 국가단위에 있어서는, 경제적 측면 뿐만 아니라, 측정이 어려운 공공적 이익까지도 분석이 가능한 새로운 평가기법이 필요하다. 넷째, 최근 과학기술의 경제사회적 영향력 증대와 함께 과학기술자집단 이외에 국민 및 정부당국의 관심과 참여가 확대되었다. 따라서 공공자금 사용의 정당성 및 성과의 제시에 대한 요구가 증대되었다.

우리나라는 국책연구개발 수행 역사가 일천한 관계로, 이에 따른 연구개발평가도 역사가 짧고, 그 수준도 낮은 상태에 머물러 있다. 따라서, 본고는 최근의 연구개발 관련 문헌에 소개된 외국 선진국들의 혁신 및 연구개발 프로그램 평가사례를 종합 분석하여, 향후 우리나라의 국책연구개발사업 평가에 도움이 될만한 교훈 및 지침을 도출

하고, 이를 정리하여 제시하기 위하여 작성하였다. 제2장에서는 외국의 12개 혁신 및 연구개발 프로그램(Innovation/R&D Programs) 평가사례를 소개하고, 제3장에서는 평가연구(Evaluation Research)이론을 기반으로, 평가목적, 기본시스템, 지원시스템, 평가결과의 활용, 그리고 평가에 대한 인식으로 구성되는 개념적 틀에 따라 외국의 프로그램 평가사례를 종합 분석하며, 제4장에서는 이러한 분석 결과를 토대로 향후 우리나라의 국책연구개발 프로그램의 평가시 도움이 될 교훈과 지침을 제시한다.

본 고에서 다루는 분석단위(Unit of Analysis)는 프로젝트가 아니고 프로그램이다. 여기서 프로그램은 구체화된 일련의 활동들을 통하여 설정된 일단의 목표들을 달성하기 위한 정부자금 또는 행정방향의 영역을 지칭하며, 프로젝트는 한 프로그램의 집행수준으로 프로그램의 목표달성에 직접적으로 공헌함으로써 하나의 최종산출물을 산출하기 위하여 자원이 활용되는 단위를 지칭한다[3]. 따라서 프로젝트의 평가와 프로그램의 평가 사이에는 평가목적, 평가방법, 평가결과의 활용 등 많은 측면에서 차이가 있다.

## 2. 외국의 혁신 및 연구개발 프로그램 평가 사례

여기에서는 미국[31], 독일, 프랑스, 스웨덴[24], 오스트레일리아[22], 영국[17], EC[19], 그리고 일본[38] 등에서 수행한 총 12개 평가사례를 통해, 외국의 혁신 및 연구개발 프로그램 평가 경험을 소개한다. 각 사례에서는 프로그램 개요와 목표, 평가목적, 평가자, 평가방법, 자료수집방법, 분석결과 그리고 평가결과의 활용 등에 초점을 맞추어 설명하고자 한다.

사례 1. 미국의 주정부 기술서비스 프로그램(The State Technical Services Program : STS)

1965년 미국 상무부(Department of Commerce)에서 시작한 프로그램으로, 첨단기술 활용에 기한 국가경제 강화, 산업기술 활용을 통한 신제품 개발 및 고용증대, 그리고 국제시장에서의 미국제품의 경쟁력 제고 등의 세가지 목표를 추구하고 있다. 이 목표를 달성하기 위하여 이미 개발된 기술을 산업에서 활용할 수 있도록, 정부가 교육, 기술지원, 그리고 정보배포를 지원한다. 프로그램의 설정은 기획 및 관리를 담당하는 지정주정부기관에서 프로그램 계획을 제안하면, 상무부 내의 STS사무국에서 승인 및 자금지원을 결정하는 형태이다.

이 프로그램에 대한 지원을 계속할 것인지 결정하기 위하여, 두번의 프로그램 평가가 실시되었다. 첫번째는 프로그램의 중요성 및 효과를 평가하여 장관에게 보고하도록 공식적으로 지정된 프로그램 평가기구인 Public Evaluation Committee에서 실시한 것으로 2년에 걸쳐 이루어졌다. 1967년에는 산업계와 학계의 전문가로 구성된 12인의 평가자가, 프로그램에 참가하는 41개 주 중 27개 주를 방문하여, 프로그램 스태프, 수혜자 및 기타 관련기관을 대상으로 심층 면담조사를 실시하였다. 1968년에는 위의 면담결과와 추가적인 현장방문 및 STS 최고경영진과의 토의를 거쳐 최종보고서를 발표했는데, 평가결과로 향후 3년 동안 프로그램 및 자금 확대, 프로그램 수혜 교육강화 및 영구적인 자문기관 설립 등을 건의하였다. 그러나 이 평가는 평가를 위한 체계적인 방법을 이용하지 않았고, 비용/편익 분석이나 다른 프로그램 혹은 대안과의 비교 등의 방법을 전혀 이용하지 않았다는 문제점을 안고 있다.

두번째 평가는, 의회에서의 예산배정을 대비하여, 지지를 확보하기 위한 정보를 제공할 목적으

로 1969년에 시행하였으며, Arthur D. Little 자문회사가 담당하였다. STS는 프로그램 평가를 위한 기초자료나 주기적 자료를 전혀 축적하고 있지 않았기 때문에, 촉박한 시간속에서 설문서와 면담을 이용하여 9개 주에서 평가기초자료를 수집하였다. 무작위 표본추출은 비용과 시간이 많이 소요되기 때문에, 적은 수의 가장 성공적인 사례에 대한 정밀평가를 실시하여 전반적인 프로그램의 성과를 추정하는 Lower Bounds Approach를 이용하였다. 성공적인 과제에 대하여 증가된 판매 수익을 측정하고, STS의 지분을 평가하며 각 주마다 첫째 및 연구효과를 계산하고, 이를 토대로 모든 주에서의 결과를 종합한 결과, 비용보다 편익이 더 큼을 발견하였다. 따라서 프로그램의 계속지원과 현장서비스의 강화를 추천하였다.

그러나, 이 프로그램은 의회에 의사결정과정에서 중단하기로 결정되었다. 프로그램 관리자들은 두번의 평가결과에 기하여 프로그램의 계속을 강력히 주장하였지만, 하원내에 프로그램을 강력하게 반대하는 실력자가 있었고, 상무부 고위층의 지원이 부족하였으며, 상무부내의 다른 부서들도 프로그램의 계속을 반대하였기 때문에, 평가결과가 무시된 사례이다.

## 사례 2. 미국 국립과학재단의 산학협동연구센터 프로그램(The NSF University-Industry Cooperative Research Centers Program : UICRC)

국립과학재단(NSF)의 산학협동연구센터는 1980년부터 대학에 설치된 산학협동연구조합으로, NSF에서 연구사업을 허가하면, 대학에서 1년간 기획을 하고, 향후 5년간 NSF에서 연구자금을 지원받는다. 이 프로그램의 목적은 연방정부차원이 아닌 주정부차원에서 연구개발투자확대를 도모하고, 연구실에서 개발된 연구결과를 산업에로의 응용·이전을 촉진하는 것이다. 대학과 산업계

의 다양한 학문적 배경을 지닌 공학자들로 구성된 공동연구팀이 연구수행의 주체가 된다. 1983년부터 체계적으로 수집한 자료의 분석이 시작되었다. 우선 NSF와의 협정에 따라 각 센터에는 내부(On-Site)평가자를 둔다. 주로 센터운영에 대한 상담역할을 수행하고 있으며, 처음에는 이들에 대하여 센터관리자들이 적대적이었으나, 시간이 지남에 따라서 그들의 역할이 센터의 효과적인 운영에 큰 도움이 된다는 것을 인식하게 되었다.

외부평가는 NSF의 Productivity Improvement Research Section에서 수행한다. 내부평가자에 의한 평가자료와, 체계적이고 범센터적인 연구설계에 의한 평가자료를 이용하여, 다양한 방법론을 적용하여 다양한 평가시점에서 평가하고 있다. 평가는 주로 센터관리자와 NSF스텝에게 지속적·건설적 피드백을 제공하기 위한 형성적(Formative)평가이다. 즉, 센터의 발전과정을 핵심사건을 중심으로 지속적·정성적으로 기술하고, 센터 관계자간 의사소통 유형에 대한 네트워크를 분석하며, 센터의 조직구조와 대학과의 관계를 기술하고, 센터 참가자의 기대와 실현성결과를 측정·비교하며, 지나온 경로와 당면한 문제점을 문서화하고, 시행착오와 성과의 피드백을 통하여 점진적으로 발전경로를 수정해 갔다. 자료수집 방법으로는 심층면담, 관찰, 문서분석, 설문서 등이 이용되며, 그 결과는 문서로 작성되어 NSF스텝, 센터관리자, 그리고 NSF 프로그램 관리자에게 유용한 정보원이 된다.

평가노력의 성공여부는 판단하기 힘들지만 가장 명확한 성공지표 중의 한가지는 평가가 프로그램의 시작부터 지속적으로 존재해 왔다는 것이다. 내부평가는 센터의 효과성에 필요하다는 인식이 확립되었으며, 작성된 많은 문서는 새로운 센터의 제안서 작성과 초기 센터의 기획 및 운영에 큰 도움이 되었다. 또한 센터의 활동과 성과에 대

하여 체계적으로 수집·분석된 정보는 센터관리자, NSF 스태프, 그리고 미래의 과제 제안자에게 유용한 의사결정 기초자료가 되었으며, 기술혁신을 위한 정책 및 프로그램 개발에 대하여, 이들이 점진적이고 실증적 태도를 취하게 도와 주었다. 특히, 내부·외부평가는 결과측면의 평가(Summative)정보보다는 프로그램 운영에 대한 과정적 정보에 관심을 둬으로써, 유용한 정보가 빠르게 발생하도록 하였으며, 평가에 대하여 좋은 인상을 심어주었다.

**사례 3. 미국의 중소기업 혁신연구 프로그램(The Small Business Innovation Research Program : SBIR)**

1970년대 중반 NSF가 실시한 실험적인 프로그램 중의 하나로, 의회와 이해관계자 집단으로부터 큰 호응을 얻어, 지속적으로 확장되어 왔다. 혁신 촉진, 중소기업을 통한 연방의 연구개발목표 달성, 혁신결과의 민간부문에서의 실용화 촉진, 등을 주요 목표로 하며, 1986년 현재 12개 연방정부 기관에서 실시하고 있다. 1982년 법령에서 몇몇 기관에 의한 프로그램의 주기적 평가를 의무적으로 규정하고 있지만, 실제로 실행되지 않았기 때문에, 1986년 법 개정을 통하여, 일반회계국(General Accounting Office : GAO)이 이 프로그램의 효과성을 평가하여 의회에 제출하도록 규정하였다.

이 프로그램의 효과성에 대한 실증적 평가는 두번 있었다. 1982년 NSF 스태프에 의한 1차 내부 평가에서는, 프로그램의 수혜 집단과 수혜기간 집단을 기술진보 정도, 기업성장 정도, 자금확보 능력에 대하여 체계적으로 비교하였으나, 유의한 차이를 발견하지 못하였고, 그 결과 또한 발표하지 않았다. 이러한 평가결과는 프로그램 확대를 결정하는 과정에서 전혀 반영되지 않았으며, 단순히

일화적이고 선별적인 증거만이 이용되었다. 1987년 일반회계국에 의한 2차 평가는, 프로그램 수혜자 특성, 프로그램이 기업의 운용과 제품에 미친 영향, 기업의 프로그램에 대한 인식, 등만을 다루었을 뿐, 프로그램의 효과성에 대한 종합적이고 핵심적인 평가는 아니었다.

이 프로그램의 평가를 저해한 요소로는 크게 두 가지를 들 수 있다. 정치적으로 입법부와 행정부에 대한 이 프로그램 이해관계자 집단의 교섭력이 강력하여, 프로그램 성과를 의심하게 할 수 있는 평가를 지원하지 못하게 하였으며, 심지어는 수혜자와 신청자에 대한 기초자료를 수집하자는 의견까지도 기각시켰다. 방법론적으로는 프로그램의 성공을 평가할 수 있는 기준이 없고, 성과가 늦게 나타나며, 성과에 영향을 미치는 프로그램 이외의 다른 영향요인들의 효과를 구별해 낼 수가 없었다. 이를 해결하기 위한 유일한 방법은 실험설계를 이용하는 것이지만, 정치적·방법론적으로 많은 어려움이 따른다. 결국 프로그램에 영향력있는 사람들은 다만 이 프로그램이 이해관계자의 필요를 만족시키고 있다는 것을 확신하고 싶었을 뿐이지, 프로그램의 효과에 대한 종합적 평가를 원치는 않았다. SBIR 프로그램은 과학적인 방법으로 검증된 효과성에 의해서가 아니고, 정치적, 상징적 필요를 만족시키기 위하여 계속 실행하고 있는 것이다.

**사례 4. 미국의 세금공제 프로그램(The R&D Tax Credit)**

이 프로그램은 1981년부터 경제회복을 위한 법령의 일환으로 실시되었으며, 특정기업에 대하여 지난 3년 동안의 평균 R&D 지출보다 증가한 부분의 25%를 세금에서 공제해 줌으로써 민간부문의 연구개발 투자를 촉진시키는 것을 목적으로 하고 있다. 법령에 의한 공식적인 평가요구는 없

었지만, 직접적인 세금공제효과에 대한 경제학자나 정치과학자들의 관심이 매우 컸으므로, 프로그램의 만기가 다가옴에 따라 그 효과를 평가하기 위한 많은 연구가 수행되었다.

첫번째 정부지원으로 1981년에 실시된 Mansfield의 연구에서는, 연구개발에 1백만달러 이상을 투자하거나, 또는 매출액이 3천5백만달러 이상되는 기업으로 1% 이상을 연구개발비로 지출하는 기업을 대상으로, 무작위 층화표본추출 방법을 이용하여 110개의 기업을 선정하였다. 각 기업의 책임자에게 세금공제효과를 추정하도록 하고, 이 결과를 이미 이 제도를 실시하고 있었던 스웨덴과 캐나다의 기업과 비교하여, 현재 상태로는 국가의 혁신활동에 큰 영향을 미치지 못한다고 결론지었다. 그러나 이 평가는 잠재적 효과를 모두 고려하지 못하여, 프로그램의 효과를 과소평가하였다는 비판을 받고 있다.

두번째 정부지원으로 실시된 Eisner, Albert, 그리고 Sullivan의 연구에서는, 어느 한 해에 연구개발에 대한 지출을 늘리면, 그 다음해의 세금공제액을 결정하기 위한 기준액도 변하기 때문에 실제 공제효과는 25% 미만이라는 주장을 제기하였다. 또한 1981년 세금회수자료를 분석하여 전체 연구개발 지출의 대부분이 고액납세 기업들에 의한 것임을 밝혀 내었다. 그리고 Standard and Poor's Compustat 자료를 분석하여, 불황일 때에는 이 프로그램이 연구개발 투자를 촉진하기 보다는 오히려 억제하는 효과를 낸다는 사실을 발견하였다. 그러나 이 연구도 단기적 효과만을 분석하였고, 프로그램 외의 다른 경제환경의 영향을 고려하지 못하였다는 반박을 받고 있다.

Baily-Lawrence-DRI 연구에서는 1982-1983년에 걸쳐 R&D 집약산업에서의 연구개발지출과, 프로그램이 없었을 때의 연구개발 지출을 추세와 계절성을 고려한 외삽법을 이용하여 추정한 금액

과 비교하여, 이 프로그램이 1980년대 연구개발 지출의 중요한 요인이었음을 주장하였다. 그러나 R&D 지출의 영향요인은 다양하므로, 인과관계를 명확히 설명하기 힘들다는 약점을 가지고 있다.

이와 같은 세금공제 프로그램에 대한 많은 연구는 의회의 큰 관심을 불러 일으켰으며, 대체로 실증적 증거는 미약하지만, 연구개발 지출을 촉진하는 데는 효과적이었다는 방향으로 의견이 모아졌다. 따라서 이 프로그램을 1988년까지 지속하기로 결정하였으며, 세금공제 비율은 20%로 재조정하였다. 많은 학자들이 세금공제 프로그램의 효과는 그 영구성에서 기인한다고 지적하였음에도 불구하고, 의회는 평가결과들이 완벽하게 일치하지 않음에 따라 일단 3년간만 더 연장하고 그 효과를 계속 평가해 보자고 결정내렸다. 그 후의 평가 결과는 이 자료에 나타나 있지 않다.

사례 5. 독일의 민간부문 연구개발에 대한 재정지원 프로그램(Financial Incentives to Promote Innovation in the Private Sector)

1979년 독일연방정부는 제조업분야 중소기업의 R&D활동을 지원하여, 산업에서의 혁신능력을 제고하고, 인력자원에 대한 투자를 증대시키고자, R&D 인력의 임금비용에 대한 보조금 지급을 실시하였다. 이 프로그램은 1987년에 종료되었으며, 후원기관이었던 FME(Federal Ministry for Economics)는 AIF(Industrial Collective Research Associations)에 평가업무를 위탁하여, 프로그램 수혜기업의 특성, 연구인력 및 다른 연구활동의 증가분, 프로그램의 효과, 프로그램의 실행 및 관리상의 효과성 등을 파악하고자 하였다.

프로그램의 단기 및 중기 영향도를 파악하기 위하여 2단계의 평가가 수행되었다. 1단계 평가에서는 1980-1981년에 걸쳐 프로그램의 시작 전과 시작 시점에서의 기업들의 현황을 평가하였다.

2단계 평가에서는 4-5년 동안의 프로그램 성과를 평가하기 위하여, 첫째로 1단계 평가결과를 기반으로 수혜기업들의 변화 정도를 조사하였고, 둘째로 프로그램이 기업의 혁신활동, 생산성 그리고 경쟁력 등에 미친 영향을 구체화하였다. 평가에 이용한 자료로는 모든 수혜기업 관련 자료, 표본 수혜기업에 대한 면담 및 설문자료, 그리고 표본 비수혜기업에 대한 면담 및 설문자료가 있으며, 자료의 분석을 위해서는 수혜기업과 비수혜기업간, 그리고 프로그램 수행 전과 수행 동안의 수혜기업의 연구개발 및 혁신활동을 비교하였다. 비교 시 주로 이용된 자료는 투입지표(R&D 스태프 수, R&D 지출), 혁신 및 생산지표, 기업의 혁신특성을 나타내는 변수들, 프로그램이 기업에 미친 정량적·정성적 효과자료, 혁신의 장애요인과 기대와의 격차, 프로그램의 관리적 측면과 관련된 자료, 그리고 다른 나라에서 수행한 비교가능한 프로그램에 대한 자료 등이다.

평가를 통하여 발견한 사실은 다음과 같다. 우선 전체기간 중 19,000개 기업 이상이 적어도 1번은 프로그램의 혜택을 받았으며, 처음 4년 동안 혜택을 받은 기업중의 반 이상이 100명 이하의 기업이었다. 프로그램 시행 첫해에는 수혜기업 중 60%가 보조금을 추가적 연구개발 및 혁신활동에 이용하였고, 마지막 해로 갈수록 비율이 점차 감소하였다. 이 비율은 캐나다와 비슷하고, 미국이나 스웨덴보다는 높은 것이다. 기업들은 보조금을 단순개발활동에서 국제적인 수준에서도 우수한 연구개발활동에까지 다양하게 이용하였는데, 주로 제품의 수정이나 개선과 관련한 한계적인 혁신에 이용하였다. 보조금 지급조건외의 지속적 변화 및 프로그램 수행기간의 불확실성은, 기업들이 보조금을 중장기 계획에 반영해야 할지를 결정하지 못하게 하여, 기업에 나쁜 영향을 끼쳤음을 발견하여, 프로그램의 효과성 제고를 위해서는 여러해

동안 지급조건을 일정하게 유지하는 것이 중요함을 알았다. 평가에 대한 저항과 기만적 답변을 배제하기 위하여는, 평가가 그들이 제공한 정보의 정확성여부를 점검하거나, 보조금의 적절한 활용 여부를 점검하려는 것이 아니라는 것을 미리 인식시켜야 함을 알았다. 이를 위해서는 프로그램의 주관기관이 아닌 외부의 독립적인 기구에서 평가를 담당하는 것이 바람직하다. 특히 기업에게 유용한 정보가 제공될 때, 평가에 대한 저항이 감소함을 알 수 있었고, 따라서 프로그램 시작부터 평가에 유용한 자료를 수집하는 것이 바람직함을 알게 되었다.

#### 사례 6. 프랑스의 기술이전 프로그램(The Regional Centres for Innovation and Technology Transfer : CRITTs)

CRITTs 프로그램은 다양한 파트너간 팀웍을 지원함으로써 중소기업사이의 기술 수요와 공급을 조화시켜, 기술이전의 실패를 방지하고자, 1981년부터 SARRT(Service d'Action Regionale de la Recherche et de la Technologie)에 의하여 시작되었으며, CRITTs에 대한 자금지원은 지방정부, 주정부 및 때때로 산업계에 의하여 이루어졌다. 프로그램이 진행됨에 따라서, SARRT는 프랑스 전역에 걸쳐 실시되고 있는 프로그램 활동에 대한 일관적인 분석을 필요로 했으며, 또한 이 프로그램은 정부기관에서 기존에 시행해 온 방법과는 많이 다른 새로운 방법이었으므로 비판의 소지가 있었다. 따라서 SARRT는 평가를 통하여 프로그램 활동의 적절성을 설명하고, CRITTs의 지위와 위상을 지원할 수 있는 자료를 중앙정부, 지방정부 그리고 다른 CRITTs에 제공하며, 끝으로 각 센터의 재정적 자립가능성에 대한 정확한 의견을 제시하고자 하였다.

평가는 Informations et Innovations Industrielles

자문회사에서 주관하여 1985년에 실시 되었다. 우선 3개의 전형적인 지역을 선택하여 프로그램에 참여한 모든 주체(파트너)들을 대상으로 면담과 설문조사를 실시하였다. 그 결과로 지역의 경제적, 기술적 현황(일반 경제자료, 연구잠재력, 지역위원회 정책 등), 기술이전 구조, 제도적 환경 및 프로그램의 향후전망 등의 내용을 포함한 평가기초자료를 확보하였다. 이 자료를 기반으로 1986년에 기술이전구조에 관한 최종보고서가 제출되었고, SARRT와의 긴밀한 토의분석이 이루어졌다. 또한 이 보고서는 1987년 연구부(Ministry of Research) 주관으로 프랑스 전지역의 프로그램 관련 인사와 산업계 주요인사가 참가한 2일간의 연구모임의 준비자료로 이용되었다.

이 평가에서는 평가기준의 설정시 CRITTs를 참가시켰고, 평가와 관련된 모든 조건들을 상세히 기술함과 동시에 평가의뢰자의 욕구를 명확히 파악하였으며, CRITTs를 합리적으로 유형화하여, 차별화된 평가를 위한 개념적 틀을 제시함으로써, 분석결과가 CRITTs 및 평가의뢰자 모두에게 잘 수용되었다. 평가가 성공적일 수 있었던 기타 이유로는, 첫째, CRITTs가 자신들의 활동의 정당성을 제시하기 위하여 평가에 호의적이었으며, 둘째, 지역과 각 센터별로 비교연구를 수행하여 센터의 연구자에게 나아가야 할 프로그램 활동의 방향을 제시해 주었으며, 셋째, 평가가 앞으로 다가올 법안의 결정과 긴밀히 연계되었기 때문에 피평가자가 평가에 적극적으로 협조했다는 것들을 들 수 있다. 이 프로그램 평가가 미친 영향은 크게 네가지로, 첫째, 중앙정부, 지방정부 및 관련 산업계에서 CRITTs의 유용성을 공식적으로 인정하게 되었으며, 둘째, CRITTs사이의 다양성 및 공통성을 인식함으로써, 센터관리자가 센터관리 및 전략수립에 필요한 분석적 틀은 확보할 수 있었고, 셋째, 센터간에 협동연구가 촉진되었으

며, 넷째, 다양한 법적·회계적 문제가 구체화되고, 해결되었다.

#### 사례 7. 스웨덴의 태양열 난방 및 펌프 프로그램 (The Programme for Solar Heating and Heat Pumps)

스웨덴은 원자력 생산체제로부터 탈퇴하기로 결정함에 따라 1975-1984년 사이에 대체에너지 원천 개발 및 실용화연구에 막대한 자금을 투입하여 총 1,200개의 연구과제를 지원하였다. 스웨덴에서는 연구개발 프로그램의 평가가 정규적인 절차로 인식되고 있고, 특히 에너지 문제에 대한 의회의 의사결정 시한이 압박했기 때문에 1985년에 프로그램 평가를 실시하였다. 평가의 목적은 1차로 에너지정책 의사결정을 위하여 정부에 정보를 제공하는 것이었고, 2차로는 정부의 기술개발 목표 달성도 점검되었다.

평가는 에너지분야의 연구개발 평가 책임기관인 Energy Research Commission에서 주관하였다. 평가위원회는 공정성과 전문성을 확보하기 위하여 독립적인 외부전문가, 평가전문가, 산업공학자 및 프로그램 수행 연구기관의 전문가로 구성하였으며, 프로그램의 자금지원 및 관리와는 무관한 기관에서 주관하였기 때문에, 평가의 독립성이 유지되었다. 평가는 2단계로 진행되었는데, 1단계에서는 탐색연구로서 외부전문가에 의한 10번의 면담과 기존 문헌고찰이 있었으며, 2단계에서는 프로그램에 의하여 잠재적으로 영향받는 모든 기관에 대한 설문조사, 프로그램에 의하여 자금을 지원받는 조직 중 10%에 대한 표본조사, 특허분석(Patent Analysis), 기술자문가와 면담, 전문가 집단의 자문, 위의 모든 자료에 대한 2명의 프로그램 관리자와의 면담 등의 6가지 자료원천을 활용하여, 연구의 과학적 수준, 관리의 효과성, 그리고 기술개발의 질적 수준 등을 평가하였다.



이 프로그램에 대한 평가는 정치적으로 중요하였고, 의회의 지대한 관심을 받았음에도 불구하고, 아무런 무리없이 진행되었다. 그 이유는 프로그램 및 연구기관에 대한 평가가 하나의 전통으로 자리잡혀 있었고, 평가수행기관이 경험이나 전문지식, 공정성에 그 인지도가 아주 높았기 때문이다. 또한 관련 정부부처의 대표자가 평가과정에 직접 참가하여 평가의 전과정을 내부에서 지켜보았으며, 그에 따라 평가결과는 관련 정부부처 및 프로그램 책임기관 모두에게 긍정적으로 수용되었다. 의회에 법안이 제출되고 난 후, 의사결정이 이루어지기 전, Royal Academy of Engineering Science가 평가보고서에 대한 토론회를 주관하였으며, 25개의 학술지에서 그 내용을 다루었다.

#### 사례 8. 오스트레일리아의 공공 연구개발 프로그램(The Public Interest IR&D Program)

이 프로그램은 오스트레일리아의 중요 산업혁신의 상업적 개발을 촉진하고, 동시에 특정산업분야의 경쟁력 제고를 위하여 필요한 공정기술개발을 촉진하고자 1978년에 시작하였다. 이 프로그램의 운영에 영향을 미치는 관련 법령이 1986년 만기됨에 따라 정부의 향후 프로그램에 대한 의사결정에 필요한 정보를 제공하고자 1985년에 상공기술부(Department of Industry, Technology and Commerce)의 의뢰로 산업경제원(Bureau of Industry Economics)에서 평가를 실시하였다. 평가목적은 프로그램의 이익을 구체화하고, 관련 분야에서의 다른 정부정책을 고려하여 이 프로그램의 계속여부를 자문하며, 또한 계속할 경우 적합한 연구과제 선정기준에 대하여 조언하는 것이었다.

평가를 위한 정보를 확보하기 위하여 기존 문서의 검토, 주정부 및 프로그램 관련 산업조직과의 토의, 그리고 각 과제의 기대이익, 실현이익

및 비용의 비교 등을 하였다. 그러나 비수혜과제에 대한 대표성있는 표본 추출이 현실적으로 불가능하여 두집단간 비교는 무산되었다. 프로그램의 총 29개 과제 중 규모, 기간 및 유형별로 다양한 19개 과제를 선정하여 정밀 검토하였다. 이중 2개는 사업화될 것이어서 기술료 수입이 발생할 것으로 예상되었고, 다른 3개는 기술적 성공을 거두어 외부효과(External Benefits)를 발생하고 있었다. 나머지 과제들 중 6개는 중도에서 포기했으며, 8개는 기술적 또는 상업적으로 실패한 것이었다.

평가팀은 연구과제 중의 일정비율은 실패가 당연하다는 것도 인정했지만, 실패한 과제중의 몇개는 더 일찍 중단함으로써 비용을 절감시킬 수 있었음을 발견하고, 과제의 중단을 위하여 좀 더 엄격한 기준을 적용할 것을 제안하였다. 평가팀은 이러한 평가결과와 프로그램의 자금지원에 영향을 끼친 사회변화를 고려하여 이 프로그램이 지속되어야 한다는 최종결정을 내렸다. 또한 앞으로는 외부효과를 발생시키는 과제에 역점을 두어야 함을 제안하였고, 이 제안은 이 평가에 이어 실시된 GIRD(Grants for Industrial Research and Development)계획의 연구과제 선정기준으로 반영되었다.

#### 사례 9. 오스트레일리아의 연구조합 지원 프로그램(Support for Research Associations : RAs)

RAs는 특정 산업분야 또는 기술분야의 기업들이 공동연구를 수행하고, 정보를 공유할 수 있도록 만든 조직으로, 1947년부터 시작되어 1989년 현재 8개의 지원금을 받는 RAs가 존재한다. 이 프로그램의 목표를 명확히 하고, RAs의 성과를 검토하며, 정부의 RAs 지원이유와 RAs의 활동을 근거로 하여 향후 역할 및 활동형태를 검토하고자 상공기술부의 의뢰로 산업경제원에서 평가

를 실시하였다.

정보수집을 위하여 설문서, 개인면담, 정부 자료분석 및 기존 보고서 조사등을 이용하였다. 최종보고서가 나오기 전에 보고서 초안이 각 면담자에게 보내져 그들의 검토 의견을 반영하였으며, 프로그램 주관 정부부처와 보고서에 대하여 공식적인 토의과정을 가졌다. 프로그램의 목표를 설정하기 위해서는 정부, 자문기관 및 RAs와 토의했으며, 기타 관련 문헌을 참고했고, RAs의 성과를 검토하기 위해서는 RAs에게 주는 혜택과 그들의 연구개발 및 기술이전으로부터 발생하는 이익을 검토하였다.

프로그램의 목표는 한번도 정부차원에서 구체적으로 명시된 적이 없어서, 프로그램 주관부처가 자의적으로 여러 위원회의 보고서 및 의도를 해석하고, RAs의 역할을 결정하였다. 따라서 평가팀은 프로그램의 목표를 RAs에 자금을 지원하여 기술이전과 연구활동을 촉진시키고, 산업의 경쟁능력을 제고하는 것이라고 규정하였다. RAs의 연구개발과제의 평가 결과, 대부분의 이익을 RAs회원이 차지함을 발견하였다. 즉, 비회원기업에 대한 서비스 제공시 경비를 부담시키거나, 경비를 부담하지 않는 기업에 대하여는 결과에 대한 접촉을 지연시켰다. 따라서 프로그램은 성공적이었고, 많은 이익을 발생시켰지만, 모두가 내부화되었다고 결론지어 지속적인 RAs에 대한 자금 지원의 정당성을 감소시켰다.

평가팀은 연구의 중복방지 입장에서 공동연구의 효과성을 인식하고, RAs의 연구활동에 기부하는 모든 기업에 대하여 연구개발 지출의 150%에 대한 세금공제 혜택을 주도록 제안하였고, 이 제안은 수용되어 오스트레일리아 연구개발 정책의 일부가 되었다. 또한 평가팀은 RAs가 기술이전 및 공동연구에 대한 긍정적인 자세 확립에서 차지하는 역할의 중요성을 인식하고, 비록 연구결

과가 전용될 가능성도 있지만, 새로운 RA의 설립과 자립할 때까지 정부가 지원하도록 건의하였으나, 기각되었다.

#### 사례 10. 영국의 Alvey 프로그램(The U.K. Alvey Programme)

1981년 일본의 제5세대 컴퓨터 개발 프로젝트 발표에 자극을 받아 영국은 정부, 학계, 산업계의 공동연구를 통하여 자국의 기술수준을 배양하며, 국가경쟁력을 증가시키고, 정보기술산업을 육성하기 위해 이 프로그램을 시작하였다. 무역산업부(Department of Trade and Industry), 국방부(Ministry of Defence), 그리고 과학공학연구심의회(Science and Engineering Research Council)의 공동후원 하에 1984년부터 향후 5년간 정보기술분야의 핵심기술을 개발함으로써 자국의 첨단산업을 육성할 목적이었다. 프로그램의 중점 전략으로는 기업간, 그리고 기업과 연구기관간 협동연구 증진, 정보기술분야 연구에 대한 정부의 지원 및 지휘 확대, 그리고 생산이나 마케팅활동 보다는 경쟁전(Pre-Competitive) 단계의 연구개발활동에의 집중 등이 있다.

Alvey의 프로그램에 대한 평가상의 특징은 첫째, 프로그램과 무관한 독립적인 외부기관에 평가를 의뢰하였고, 둘째, 프로그램의 초기부터 그 운영과 결과를 평가하는 것의 중요성을 깊이 인식하였기 때문에, 프로그램의 수행기간 동안 평가작업이 지속적으로 이루어지도록 하는 실시간(Real Time) 평가를 실시하여, 평가결과의 피드백을 통해 프로그램수행을 돕고자 하였다는 것이다. 다음의 세 집단에게 연구를 의뢰하였다.

① PREST(The Programme of Research on Engineering Science and Technology, University of Manchester) : 프로그램의 구조와 조직 평가

② SPRU(The Science Policy Research

Unit, University of Sussex) : 프로그램의 영향도 평가(프로그램의 적합성 평가, 기술적·경제적 목표 달성도 평가)

③ CBS(The Center for Business Strategy, London Business School) : 영국 정보산업 각 분야에 미치는 영향을 평가(중간 탈퇴)

특히, 실시간(Real Time) 평가를 시행한 것은 사전(Ex-Ante)평가나 사후(Ex-Post)평가 시행보다 여러가지 장점이 있는데 첫째, 프로그램 종료 후에는 얻을 수 없는 자료와 전문가의 의견 수집이 가능하고, 둘째, 평가결과를 프로그램 책임자에게 피드백하여 프로그램의 성과나 지휘에 영향을 줄 수 있으며, 셋째, 평가욕구에 맞는 자료를 가장 적절한 시점에 능동적으로 확보할 수 있다는 점 등이 그것이다.

프로그램 평가의 대상으로는, 프로그램의 실행에 관련된 문제, 프로그램의 추진도와 영향에 관한 문제, 프로그램의 적절성에 관한 문제 등을 다루었다. 평가방법의 선정시 프로그램의 시작 이유, 전반적인 목표 및 세부 목표, 그리고 목표달성을 위한 전략 등을 고려하여 다양한 접근방법을 택하였다. 예를 들면, 프로그램의 적합성 평가를 위해서는 설문조사, 구조화된 면담, 역사적 분석, 국제 비교 등을 이용하였으며, 기술적·경제적 영향도 평가를 위해서는 설문조사 및 인터뷰, 특허분석, 계량서지학적 분석, 사례분석 등을 이용하였다.

예컨대 1986-1987년에 실시된 Alvey 프로그램의 4개 기술분야 중의 하나인 VLSI 프로그램의 평가에서는 프로그램이 학술분야의 연구개발 활동에 미친 영향도를 분석하였다. 평가를 통해 프로그램이 미친 영향도, 연구결과의 활용도, 프로그램의 진척도, 연구 장애요인, 산학 공동연구에 미친 영향, 공동연구의 효과, VLSI 목표 및 전략의 적합성들을 알아내었다. 평가의 오류를 방

지하기 위해서 인터뷰나 설문조사시 프로그램에 대하여 비우호적인 집단을 포함시켰고, 다른 기술분야와의 비교를 통한 상대적 성과를 도출하였으며, 편의(Bias)가 예상되는 항목에 대한 중복측정을 시도하였다. 또한 이러한 실시간 평가자료는 프로그램 종료 이후에 수행한 최종평가시에 중요한 평가자료를 이용하였다.

사례 11. EC의 ESPRIT 프로그램(The European Strategic Programme for R&D in Information Technology)

ESPRIT는 정보기술분야에서 유럽 최대규모의 프로그램이며, 최초의 산학연 협동연구체제이다. 1983년 개념정립 단계를 거쳐, 10년 동안 2단계로 프로그램이 수행되었으며, Microelectronics, Information Processing System, 그리고 Information Technology Application Technology 등의 분야에서 연구개발을 수행하였다. ESPRIT의 목표는 크게 세 가지로, 첫째, 정보기술분야에서 경쟁전(Pre-Competitive)단계 연구개발 협동체 구축, 둘째, 유럽 정보기술산업의 경쟁력 확보에 필요한 기반기술 제공, 셋째, 정보기술분야에서 국제적 표준을 확립하는데 필요한 준비를 하는 것 등이다.

ESPRIT에서는 연구과제 수준에서 정기적인 평가를 매년 시행하고(Ex-Ante), 프로그램 수준에서는 2년에 1번씩(Interim) 수행한다. 1985년 "Mid-Term Review of ESPRIT"를 시작으로, 1986년 "ESPRIT I Progress and Results," 2단계 ESPRIT를 위한 준비작업으로 "Broadening Operation," 그리고 1987년 첫번째 사후평가인 "Final assessment of Microelectronics Programme"을 실시하였다. 즉, 지속되는 프로그램에 대한 체계적·정기적 점검뿐만 아니라, 연구개발 활동의 방향수정, 연구과제의 계속 및 중단결

정 등을 위한 평가를 동시에 맞게 수행하고 있다.

1985년 지금까지 프로그램이 원래의 목적에 맞게 수행되었는가를 파악하고, 이에 대한 정보를 2단계 ESPRIT 기획시 반영하기 위하여 Mid-Term Review of ESPRIT를 실시하였다. Panneborg박사 팀이 맡아서 수행했으며, 131개 조직에서 477명에게 설문조사하였다. 그 결과 프로그램은 잘 설계되었고, 원래의 목적을 달성하고 있으며, 초기의 기대보다 빨리 추진되고 있다고 결론지었다. 또한 향후 프로그램의 발전을 위하여 경쟁전 단계 R&D에 역점을 두어야 하고, 기술통합과제에 관심을 기울여야 한다고 제안하였다. 이러한 제안은 EC위원회의 의사결정에 반영되었다. 1986년 Mid-Term Review에서 다루지 않았던 구체적인 기술결과와 개별과제의 영향도 및 산업에의 응용도를 측정하기 위하여 ESPRIT I Progress and Result를 수행하였다. 정보기술분야 이사회 내의 평가조직에서 담당했으며, 6개월마다 제출되는 각 과제에 대한 외부기술평가자의 중간평가보고서를 기초자료로 이용하였다. Broadening Operation에서는 2단계 ESPRIT에 포함시켜야 할 기술을 결정하기 위하여 회원국의 59개 조직으로부터 자료를 수집하였다. 정보기술분야 기업대표로 구성된 Panel에서 평가에 포함되어야 하는 주제에 대한 초기제안서를 작성하고, 정보기술분야 산업 및 관련기관의 대표자가 참가하는 일련의 워크숍을 개최하여 초기제안서의 기술범위 및 내용에 대한 토의를 하여, 평가조직의 의견을 많이 반영하였다. 또한 이미 발표된 2가지 중간평가 결과를 참고하여, ESPRIT I의 단점을 최대한 극복하려고 노력하였다. 1987년 Microelectronics 프로그램이 끝났을 때, Van Overstraeten이 이끄는 독립적인 외부전문가팀이 프로그램의 실행과 관리적 측면, 프로그램 성과측면, 그리고 지역사회에의 기여도 측면에서 평가하였다. 우

편 설문조사와 참여기업 대표와의 면담을 이용하였으며, 특히 면담결과의 객관성 제고를 위해서는 정보기술분야 이사회내 평가조직의 1인이 면담과정에 함께 참가하여 개별적으로 보고서를 작성하여 비교하였다. 평가결과 이 프로그램은 전반적으로 성공적이었고, CAD 및 설비분야에서 신기술 및 제품개발에 기여했으며, CAD학회 및 설계도구관련 기술발전에 공헌한 반면, 자원제약으로 인하여 Microelectronics 산업의 국제경쟁력 제고에는 큰 영향을 끼치지 못했음을 발견하였다.

이러한 중간평가와 종료평가 과정을 거쳐, 정보기술분야 이사회에서는 평가가 기획 및 의사결정을 위한 필수요소임을 인식하고, 운영부서외에 독립적인 내부평가부서를 만들어 일상적인 관리 및 통제업무를 수행하도록 하였다. 이 영구적인 평가집단은 지속되는 프로그램에 대한 체계적인 정기적 점검과 함께, 프로그램의 방향설정과 계속 및 중단결정을 위한 중간·종료평가를 수행하는 외부평가집단에게 평가관련 정보를 제공한다. ESPRIT II에서는 지금까지 계량적 목표가 없었기 때문에 평가가 정성적으로만 실시된 점을 감안하여 기술적 목표를 최대한 계량화하려고 노력하였으며, 외부평가를 위한 정보와 자료를 프로그램 시작부터 수집하기로 결정하였다.

#### 사례 12. 일본의 대형공업기술 연구개발 프로그램

대형공업기술 연구개발 프로그램은 통산성 산하의 공업기술원에서 1966년에 설립한 프로그램으로, 국민경제상으로 긴급하고 선도적인 대규모 공업기술로서 연구개발에 거액의 자금과 매우 긴 시간이 필요하며, 매우 큰 위험부담을 수반하기 때문에, 민간부문의 지속적인 연구개발에 의해서는 수행할 수 없는 대규모 산업기술에 대해 정부가 지원함으로써, 정부의 자금과 민간부문의 R&D 능력을 결합하여 국가기술수준을 향상시키

는 것을 목적으로 하고 있다. 프로그램 초기의 과제 선정 기준은 수출 증대를 위한 광공업 분야의 국제 경쟁력 강화와 사회발전에 주된 초점이 주어졌으나, 1975년 산업기술심의회 의 건의에 따라 다소 수정되었다. 즉 1970년대에 들어서서 계속되는 무역흑자로 인해 수출 장려정책은 필요없게 되었고, 그 보다는 환경 공해의 문제가 크게 제기 되었으며, 축적된 과제관리 경험은 효율적인 관리를 중요시하여 연구목표나 계획을 중요시하게 하였다. 또한 사회경제적 요구도 많이 반영하였다. 그 예로 1차 석유위기가 다가왔을 때 대체에너지 개발 및 에너지 절감을 위해 에너지관련 기술개발에 적합한 새로운 연구개발 제도를 검토했으며, 1980년대에는 기초연구의 중요성을 인식하게 되었다.

이러한 프로그램에 대한 일본에서의 평가 특징은, 프로그램 관리자가 직접 평가에 참여하고, 정부와 다양한 프로그램 자문기구들간의 조정과 합의에 의해서 이루어지는 내부평가라는 것이다. 이는 서구 국가에 비해서 독립적이고 객관적인 평가를 할만한 전문평가자가 부족한 현실과, 공식적으로 다른 사람을 비판하는 것을 싫어하고, 논쟁을 피하려는 특성이 제3자에 의한 독립적인 평가를 거부하기 때문이다. 따라서 연구과제 평가를 위한 소위원회가 있음에도 불구하고 과제의 사전 평가와 중간평가는 공업기술원 사무국의 내부활동으로만 이루어지고, 평가소위원회는 과제의 사후평가에만 관여한다. 사후평가도 특정 업무에 대한 검토 및 점검이라는 협의의 평가보다는 과제의 지원과 촉진, R&D 정책 수립과 개선을 위한 광의의 평가가 중심이 된다. 이는 성과를 객관적으로 평가할만한 전문가가 부족하다는 것을 인식하고 있고, 동시에 미국의 산업기술변화에 신속히 대응하는 것을 중요시하기 때문이다.

그러나 이러한 합의추구 평가제도는, 결과의 수

용과 실행이 용이하다는 장점이 있는 반면, 문제점도 지속적으로 제기하고 있다. 합의를 통한 내부평가는 프로그램의 점진적 개선 및 신규 프로그램의 시작에는 긍정적인 효과가 있지만, 프로그램의 폐기 기능은 제대로 수행할 수 없다는 것과, 정부관료에 의하여 시작된 내부평가 및 합의추구 과정은 공정성의 문제가 있다는 것 등이며, 따라서 1980년대 초반부터 평가제도의 변화가 필요하다는 인식이 크게 대두되어, 혁신정책을 검토하는 총리자문기구의 보고서 등에서 제도의 개혁을 요구하고 있다. 또한, 예산적자로 인해 이해관계자간 투쟁이 증대되었고, 연구의 방향이 기초연구 중심으로 이전함에 따라 의사결정자에게 더 많은 전문지식이 필요하다는 것 등이 합의추구 방법의 한계점 및 장애물로 등장하고 있다. 그러나 기존의 제도에 깊이 뿌리 박힌 문화와 전문평가자의 부족이라는 두가지 큰 문제가 이러한 개혁이 단기적으로는 불가능하고 장기적으로 이루어져야 한다는 결론을 유도하고 있다.

위의 12가지 사례에 대하여 각각의 장·단점과 우리나라 상황에서의 적용가능성을 개략적으로 정리해 보면 다음과 같다. 사례 1의 경우 촉박한 시간과 적은 예산을 고려한다면 Lower Bounds Approach는 프로그램의 효과성을 파악할 수 있는 적절한 비용/편익분석 방법 중의 하나이다. 그러나 Public Evaluation Committee에서 심층면담조사를 중심으로 수행한 정성적 평가는 평가결과의 신뢰성 확보에 큰 약점을 가지고 있다. 따라서 사례 11에서와 같이 정성적 평가방법의 약점을 개선하고자 하는 노력이 추가되어야 할 것이다. 또한 정책결정자를 비롯한 평가결과 활용자 집단의 요구를 충족시키지 못하고, 이들의 관심과 평가과정의 참여를 유도하지 못함으로써, 평가결과가 의사결정 과정에 전혀 이용되지 못한 것을 알 수 있다. 사례 2에서 내부평가자는 센터의

운영에 대하여 상담역할을 수행하여 센터의 효과적인 운영에 기여하고 있으며, 외부평가자도 과제 제안자, 센터관리자 및 프로그램 관리자에게 건설적인 정보를 지속적으로 제공함으로써 평가에 대한 인식을 제고시키고, 따라서 유용한 평가정보의 발생을 촉진하고 있다. 그러나 이러한 평가는 프로그램의 운영과 관리 측면에서는 매우 유용하지만, 프로그램의 전반적인 효과성에 대한 정보를 제공하지 못하므로, 프로그램의 중단과 계속 결정에는 부적합하다. 사례 3에서 NSF는 프로그램의 효과성을 파악하기 위하여 프로그램 수혜집단과, 수혜기간집단 중 다른 프로그램으로부터 자금을 지원받은 집단의 성과를 비교하는 체계적인 준실험설계 방법을 이용하고 있다. 반면에 일반회계국에서 수행한 평가는 프로그램의 현황을 기술하는 프로그램 모니터링에 가깝기 때문에 의사결정을 위한 유용한 정보는 제공하지 못하고 있다. 이 사례는 정치적 영향력에 의하여 평가의 수행과 평가결과의 활용이 어떻게 제한받는지 잘 보여주고 있다. 사례 4의 첫번째 연구에서는 각 기업의 책임자에게 프로그램의 효과를 추정하게 하고, 그 결과를 유사한 프로그램을 수행하고 있는 다른 나라의 결과와 비교함으로써 평가결과의 객관성을 높이고 있고, 두번째 연구에서는 세금회수자료와 기존의 방대한 통계자료를 분석하여, 프로그램의 실제적인 효과를 추정하고 있으며, 세번째 연구에서는 R&D 지출에 대한 시계열 자료를 추세와 계절성을 고려한 외삽법을 이용하여 프로그램의 효과를 분석하고 있다. 첫번째 연구의 경우 프로그램의 잠재적인 효과를 고려하지 못하고 있기 때문에 프로그램 효과를 과소평가한다는 지적을 받고 있으며, 두번째와 세번째 연구에서는 단기적 효과만을 고려하고 있고, 다른 경제환경의 영향을 고려하고 있지 못함을 지적받고 있다. 이는 비교 집단 없이, 동일한 수혜기업에 대하여 프로그램

시작 전·후의 상태를 비교함으로써 프로그램의 외부 영향요인을 적절히 통제하지 못하기 때문이다. 사례 5에서는 비교집단이 있는 준실험설계 방법과 시계열 자료분석을 동시에 이용하고 있다. 즉 연구개발 및 혁신활동에 대하여 프로그램 수혜기업과 비수혜기업을 비교하고, 동시에 수혜기업에 대해서는 프로그램 시작 전·후의 변화를 측정함으로써 평가결과의 객관성 및 신뢰성을 제고하고 있다. 사례 4와 5에서와 같이 시계열 자료의 분석을 위해서는 프로그램 시작전부터 프로그램과 관련된 환경이나 프로그램 참가집단에 대한 방대한 양의 기초자료 수집이 전제되어야 한다. 즉 국가적 차원에서 평가기초자료를 수집하거나, 프로그램 내부의 평가집단에서 평가관련 자료를 체계적으로 수집해야만 한다. 사례 6에서는 인지도가 높은 평가전문기관이 평가를 수행하고 있으며, 평가기준의 설정시 각 센터관리자를 참가시키고, 평가주관기관이 평가의뢰자의 요구를 명확히 파악하고 있다. 또한 지역별, 센터별 비교연구를 수행하여, 센터 특성에 따른 합리적인 유형화 및 차별화된 분석 틀을 제공함으로써 평가결과의 수용도를 제고시키고 있다. 사례 7에서는 프로그램의 자금지원 및 관리와는 무관한 인지도가 높은 평가책임기관에서 평가를 수행하여 평가의 독립성과 공정성을 높이고 있으며, 프로그램의 영향을 받는 집단에 대한 다양한 자료를 비교·종합함으로써 평가결과의 신뢰성을 높이고 있다. 특히 평가에 대한 오랜 전통은 평가의 수행 및 결과의 수용에 큰 도움이 되고 있음을 볼 수 있다. 사례 8에서는 프로그램의 대표성 있는 과제를 표본추출하여, 기대이익, 실현이익 및 투입된 비용을 비교하고, 외부효과의 발생여부를 검토하고 있다. 그러나 비수혜과제에 대한 대표성 있는 표본추출의 어려움으로 인하여 원래 의도했던 통제집단의 의한 비교분석은 수행하지 못하고 있다. 이것은

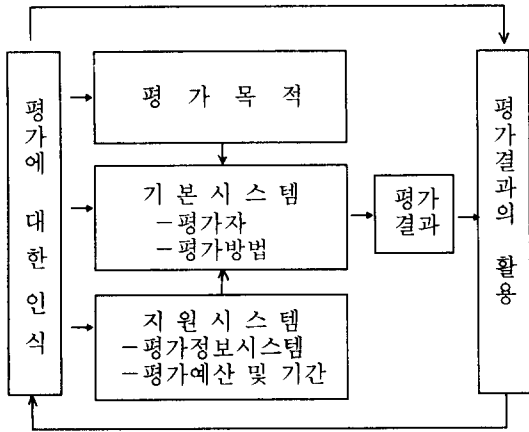
사후적으로 비교집단을 설정하는 준실험설계의 가장 큰 문제점인데, 프로그램의 기획시 평가설계를 병행함으로써 문제를 극복할 수 있다. 사례 9에서도 경제적인 접근방법을 이용하여, 연구조합에게 제공하는 혜택과 그들의 연구개발 및 기술 이전으로부터 발생하는 이익을 비교하고 있으며, 평가를 통하여 지금까지 모호했던 프로그램의 목표를 명확히 정의하고 있다. 사례 10에서는 실시간 평가를 수행함으로써 사전평가나 사후평가가 갖지 못하는 많은 이점을 얻고 있다. 또한 평가내용에 따라 다른 연구집단 및 접근방법을 도입하여 평가의 전문화 및 평가결과의 정확도를 향상시키고 있으며, 평가의 오류를 극소화하기 위하여 다양한 방법을 적용하고 있다. 그러나 실시간 평가의 경우, 요구되는 시간과 인력이 막대하기 때문에 비용이 크다는 문제점을 안고 있다. 사례 11에서는 2년에 한번씩 정기적으로 프로그램 수준에서 중간목표 달성도를 점검하고, 향후 프로그램의 기획시 수정·보완되어야 할 내용을 제안하는 중간평가를 실시하고, 프로그램이 종료되면 다양한 측면에서 성과를 측정하는 사후평가를 실시한다. 또한 프로그램 내부에 독립적인 평가부서를 설치하여 일상적인 프로그램 관리 및 통제업무와 함께 평가기초자료 수집을 통한 외부평가 지원업무를 수행하도록 한다. 사례 12에서는 프로그램 관리자가 평가과정에 직접 참여하고, 정부와 다양한 프로그램 자문기구 사이의 조정과 합의를 거쳐 평가가 수행된다. 이러한 내부평가는 평가결과의 수용과 실행을 촉진하고, 프로그램의 점진적인 개선과 신규 프로그램의 시작에는 도움이 된다는 장점이 있는 반면, 프로그램의 중단결정을 하지 못하며, 평가의 객관성과 공정성이 낮아진다는 단점이 있다. 위의 사례에서 보면, 평가의 독립성과 공정성이 중요하기 때문에 문화적 전통이 특이한 일본을 제외하고는 사례에 나와 있는 모든 국가

에서 외부평가를 이용하고 있음을 알 수 있다.

우리나라와 같이 전국가적 차원에서 연구인력과 연구비 규모가 작고, 연구개발의 효율성을 제고하고자 중복투자를 기피하는 나라에서는 수행하는 연구개발과제의 종류와 수가 적기때문에 비교가능한 통제집단을 설정하는 것이 사례 7에서와 같이 사실상 어렵다. 따라서 사례 3과 5에서와 같은 통제집단에 의한 준실험설계 방법보다는 비교집단이 필요없는 분석방법이 더 타당할 것이다. 프로그램 참가집단이나 프로그램과 관련된 경제 환경에 대한 시계열자료가 적절히 축적되어 있다면 프로그램 시작 전·후의 상태를 비교하거나 추세를 분석하는 시계열 자료 분석이 사례 4와 5에서와 같이 가능할 것이고, 프로그램의 성과가 화폐가치로 나타내어질 수 있는 경우에는 사례 1, 8 그리고 9에서와 같이 비용/편익분석 방법이 적용가능할 것이다. 또한 모든 사례에서 필수적으로 나타나고 있는 심층 면담이나 설문조사는 충분히 가능하며, 이러한 방법을 적용할 때에는 사례 7, 9, 10 그리고 11에서와 같이 평가결과의 객관성과 신뢰도를 높이려는 노력이 병행되어야 할 것이다.

### 3. 사례분석 및 토의

우선 지금까지 소개한 외국의 사례들을 비교·분석·종합하는 데에는 기준이 될만한 개념적 틀이 필요하다. 이 논문에서는 프로그램 평가활동의 구성요소들을 평가목적, 기본시스템, 지원시스템, 평가결과의 활용 및 평가에 대한 인식으로 나누고[6], 이에 따라서 사례들을 분석하고 사회과학 분야에서 발전된 평가이론과 접목시키고자 한다. 전체적인 구성요소 및 흐름관계는 <그림 1>에 나타나 있다.



〈그림 1〉 프로그램 평가를 위한 개념적 틀

3.1. 평가목적

프로그램 평가는 목적과 시기에 따라서, 프로그램의 수행전에 프로그램 계획의 타당성을 증진시키거나 프로그램의 수행 중에 프로그램의 관리 및 운영을 개선하고자 수행하는 형성적 평가와, 프로그램의 종료 후 정책 결정자나 이해관계자의 정보 요구를 충족시키기 위해 수행되는 총괄적 평가로 나눌 수 있고[34], 평가의 초점에 따라서 프로그램의 개념화 및 설계를 위한 평가, 프로그램 실행의 모니터링, 그리고 프로그램의 영향도와 효율성을 포함하는 효용성 평가 등으로 나눌 수 있다[32]. 이러한 평가의 분류는 사전(Ex-Ante), 중간(Interim), 사후(Ex-Post)평가의 분류와 유사하다. 사례 2, 6 그리고 12에서는 프로그램 모니터링과 형성적 평가에, 사례 1, 3, 4, 5, 7, 8 그리고 9에서는 프로그램 효용성 평가에, 그리고 사례 10과 11에서는 둘 모두에 초점을 맞추고 있다. 특히 Gibbons[15]는 모니터링과 평가를 명확히 구분하여, 전자는 일어난 사건 및 활동에 대한 단순한 확인 및 측정, 그리고 후자는 프로그램의 효용성에 대한 판단을 내리는 것이라고 정의하고, 지금까지는 판단을 위한 평가보

다도 단순한 모니터링 위주로 평가가 진행되어 왔음을 지적하고 있다. 프로그램의 효과성이나 효율성에 대한 정보가 확보되어야 프로그램의 계속 지원이나 중단, 또는 확대, 축소 및 수정의 의사 결정을 내릴 수가 있다. 이러한 프로그램 영향도 평가결과의 이해와 해석을 위해서 반드시 선행되어야 하는 것이 모니터링이다. 프로그램의 계획이 아무리 합리적이고 바람직하다 할지라도 프로그램의 실행이 원래의 계획대로 이루어지지 않으면 의도한 효과를 가져올 수 없을 뿐만 아니라, 향후 관련 프로그램의 도입이나 기존 프로그램의 개선에 아무런 도움을 줄 수 없게 된다[12,29].

모니터링을 위해서는 관찰, 프로그램 기록 자료 분석, 프로그램 관리자나 스텝 또는 프로그램 참가자에 대한 인터뷰나 설문조사와 같은 정성적인 방법이 주로 이용되며, 프로그램의 효율성, 효과성 또는 영향도 평가를 위해서는 앞의 정성적인 방법외에 진실험설계, 준실험설계 또는 비용/편익 분석 등의 정량적인 방법을 추가하여 평가 결과의 신뢰도 및 통계적 유의성을 제고할 수 있다.

평가의 기획 단계에서 반드시 짚고 넘어가야 하는 문제 중의 하나는 평가를 위해 선택된 평가 척도의 타당성문제가 발생하지 않도록 명확하고 구체적으로 프로그램의 목표를 정의해야 한다는 것이다[40]. 또한 앞에서도 말한 바와 같이 프로그램의 설계와 활동을 정확히 파악해야만 프로그램의 관리를 잘못된 것인지, 아니면 프로그램의 효과가 없는 것인지를 정확하게 알 수 있다[43]. 이러한 명확화를 위해서 이용할 수 있는 방법이 평가성조사(Evaluability Assessment)이다. Rutman[33]은 프로그램 관련 문서조사, 프로그램 관리자 및 스텝과의 면담, 그리고 프로그램이 실행되고 있는 현장을 방문조사하여 얻은 결과를 종합하여, 평가대상이 될 목표와 구성요소, 이들간의 인과관계를 명시함으로써, 평가가능한 프로그



램의 모형을 개발할 것을 제안하고 있다. 이러한 과정을 거쳐 모호하거나 프로그램 실시를 위해서 크게 과장되었던 프로그램 목표를 현실적으로 재조정하고, 평가문제를 명확히 하며, 평가결과와 활용을 촉진하고, 부수적으로 프로그램 관리자에게 프로그램 설계 및 전달과정에서의 문제점을 지적해 줄 수 있다.

### 3.2. 기본시스템

#### 3.2.1 평가자

평가자의 소속에 따라서 평가는 자체평가, 내부평가 그리고 외부평가로 나뉜다[2]. 자체평가는 프로그램의 집행을 담당하고 있는 사람들 자신이 수행하는 평가인데 신뢰하기 어렵다. 그 이유는 평가의 기본가정이 프로그램은 효과가 없다는 것임에 비해 관리자의 원리는 정반대이기 때문이고, 일상적인 업무속에서 평가에 할애할 시간이 적고, 평가에 대한 전문지식도 없기 때문이다[32]. 반면에 평가는 사회적 과정이며, 동시에 학습과정임을 인식하고, 자체평가는 프로그램 집행자의 전문성 증진에 필수적이라는 시각도 있다[19]. 평가를 제도화하는 데에는 저항이 필연적이고, 따라서 저항이 내재되어 있는 부적절한 평가기준을 무리하게 적용하기 보다는 평가대상집단과 긴밀하게 협의하는 자세를 취하는 것이 자체평가문화의 정착에 필요하다[7].

사례 12를 제외하고는 모든 사례에서 외부평가만, 또는 내부평가와 외부평가를 동시에 수행하고 있음을 알 수 있다. 이는 정부프로그램 또는 범국가적 프로그램의 특성상 평가의 객관성과 공정성이 제일 중요하게 여겨지기 때문일 것이다. 내부평가의 장점은, 사례 2, 10, 11 그리고 12에서 나타나듯이, 평가자가 프로그램의 실행과 근접하게 위치해 있기 때문에 프로그램의 특성파악 및 자

료수집이 용이하고, 프로그램의 효과적 운영과 점진적 개선 및 신규 프로그램의 도입에 도움이 된다는 것이다. 반면에 단점은 객관성과 신뢰도가 부족하고, 프로그램의 중단결정을 내리기 힘들며, 또한 평가에 대한 전문지식이 부족할 수 있다는 것이다. 따라서 내부평가만을 수행하는 것은 바람직하지 않고, 사례 2, 10 그리고 11처럼 외부평가를 내부평가가 지원하도록 하거나, 아니면 다른 경우들처럼 외부평가만을 실시하는 것이 더 좋은 방법일 것이다.

내부평가로 하여금 외부평가를 지원하도록 함으로써 얻을 수 있는 가장 큰 장점중의 하나는 평가에 필요한 기초자료를 프로그램의 시작부터 수집함으로써, 평가요구에 맞게 다양한 방법과 다양한 자료원천을 이용하여 평가결과와 신뢰도와 정확도를 최대한 높일 수 있다는 것이다. 사례 6과 7에서 보면 전문성과 인지도가 높은 독립적인 외부평가기관을 선정함으로써, 피평가자의 저항을 최대한 줄이고, 평가결과와 수용도를 높일 수 있음을 알 수 있다.

사례 10. 에서 보면 평가문제에 따라서 가장 적합한 평가집단을 다르게 지정하는 것을 볼 수 있고, 사례 4에서는 동일한 연구문제에 대하여 다양한 평가집단이 다양한 평가방법을 적용함으로써 상호비교를 통해 평가결과와 신뢰도 및 정확도를 높임을 볼 수 있다. Ormala[26]는 덴마크, 핀란드, 노르웨이 그리고 스웨덴의 연구개발 평가경험을 토대로, 평가의 신뢰도를 높이기 위해서는 외부평가의 경우 평가자의 공정성과 전문성이 중요하고, 내부평가의 경우는 합의된 원칙이나 규칙을 따르고, 동시에 평가절차를 공개하는 것이 바람직하다고 제안하고 있다.

#### 3.2.2 평가방법

프로그램 평가에 이용되는 방법은 크게 정성적

인 방법과 정량적인 방법으로 나눌 수 있다. 사례에서 살펴보면 거의 모든 경우에서 정성적인 방법은 필수적으로 이용하고 있고, 여기에 추가적으로 몇몇 정량적인 방법을 이용하고 있음을 알 수 있다. 많이 이용되는 정성적인 방법으로는 프로그램관련 문서 및 기록 분석, 관찰법, 사례연구, 인터뷰, 설문조사, 청문회, 역사적 자료 분석, 국제 비교 등이 있고, 정량적인 방법으로는 계량서지학적 분석, 계량경제학적 분석, 비용/편익 분석, 그리고 사후측정 비교집단 설계, 단일집단 사전사후 측정설계 및 시계열자료분석 등과 같은 진실험설계방법 등이 있다. 정성적인 방법과 정량적인 방법에는 각각의 장단점이 있다[28]. 정성적인 방법은 적은 수의 평가대상에 대하여 깊고 자세한 분석을 가능하게 하고, 분석방법에 의해 크게 제약받지 않는다. 또한 사람과 상황에 가깝게 접근하기 때문에, 실제로 일어나는 일을 파악하여 현상의 원리와 인과과정에 대한 이해가 가능하게 해준다. 반면에 정성적인 방법은 비용이 많이 들고, 오해의 위험도 크며, 사용의 어려움도 많다. 정량적인 방법은 표준화된 자료를 이용하므로 광범위한 분야의 자료에 대한 비교 및 통계적 통합이 용이하고, 따라서 결과의 신뢰도와 일반화 가능성을 높여준다. 하지만 정량화는 현장을 경시하는 경향이 있으며, 본질적인 문제와는 멀어질 가능성이 크다. 이러한 논쟁속에서 나타나고 있는 합의점은 두 방법 모두 가치가 있고, 정당하다는 결론이다[27]. 정성적인 방법은 프로그램의 설계와 모니터링에, 정량적인 방법은 영향도 평가와 효율성 분석 등에 적합하다. 평가과정에서 가능한 한 다양한 방법을 이용하고, 서로 비교함으로써, 결과의 타당성을 제고하고, 여러가지 측정오류를 방지하며, 평가문제와 의사결정자의 요구에 따라서 가장 적합한 방법과 자료를 이용할 것을 제안하고 있다[31].

사례에서 보면 정성적인 방법의 단점을 극복하기 위한 노력이 많이 이루어지고 있음을 알 수 있다. 사례 10에서는 인터뷰와 설문조사시 프로그램에 대하여 비우호적인 집단(예를 들면, 연구과제를 신청했는데 기각된 사람)을 포함시켰고, 편익(Bias)가 예상되는 항목에 대해서는 여러사람에게 중복 측정하였으며, 성과의 판단시(절대적인 기준이 아닌) 비교를 통한 상대적인 기준에 의하여 의사결정을 내렸다. 사례 11에서는 면담결과의 신뢰성을 제고하고자, 면담과정에 외부평가자 1인과 내부평가자 1인이 참석하여 각기 다른 면담결과보고서를 작성하고 서로 비교하였다. 사례 7에서는 설문조사, 특허자료, 면담결과, 등 총 6가지 자료원천을 종합하여 연구의 과학적 수준, 프로젝트 관리의 효과성, 그리고 기술의 질적수준 등을 평가하고 있다. 사례 9에서는 보고서 초안을 모든 면담자 및 관련 정부부처에 미리 배포하여 자료의 정확성여부를 피드백 받음으로써 최종보고서의 신뢰도를 높이고 있으며, 사례 6과 7에서는 최종보고서에 대한 발표회 및 토론회를 개최하고 있음을 알 수 있다.

프로그램이 의도한 효과를 나타내었는가를 평가하는 데에 가장 신뢰할 수 있는 방법은 진실험설계이다. 하지만 현실적으로 이를 적용하기는 어렵는데, 첫째, 진실험설계를 하려면 프로그램의 기획시에 함께 평가도 설계해야 하는데, 평가는 프로그램의 진행중이나, 완료되었을 때 이루어지는 경우가 많으며, 둘째, 진실험설계를 하는 데에는 시간, 자원 및 정치적·도덕적인 다양한 제약이 있기 때문이다. 이의 대안으로 적용가능한 것은 진실험설계이다. 사례 11에서와 같이 프로그램 수혜집단과 기각집단을 쌍대비교하는 사후측정 비교집단설계, 사례 5에서와 같이 프로그램 참가기업의 프로그램 참가 전후의 상태를 비교하는 단일집단 사전사후측정설계, 그리고 사례 4에서

보듯이 연구개발 지출의 추세와 계절성을 고려한 외삽법 등을 이용하는 시계열자료 분석 등이 이용되고 있음을 알 수 있다. 목표집단의 80% 이상이 프로그램에 참가하고 있으면 비교집단을 형성할 수 없으므로 재귀적 통제에 의한 평가설계가 가능하다. 하지만 그렇지 않은 경우에는 비교집단을 이용한 실험설계가 재귀적 통제에 의한 설계보다 더 바람직한데, 이는 대부분의 경우 재귀적 통제에 의한 영향추정치의 신뢰도가 낮기 때문이다[32]. 또한 재귀적 통제에 의한 평가설계라도 프로그램 실시 전후에 측정이 더 많을수록 추정치의 신뢰도가 더 높아진다[21]. 하지만 준실험설계를 현실적으로 실시하는 데에도 많은 어려움이 따르므로 사례 8에서와 같이 중도에서 포기하게 되는 경우도 발생하게 된다. 신뢰성있는 결과를 얻기 위한 실험설계를 프로그램평가에 적용하기 위해서는, 첫째, 프로그램의 설계시부터 평가의 중요성을 인식하고, 가장 적합한 평가설계를 프로그램 설계와 함께 고려해야 할 것이며, 둘째, 프로그램 실행 전후에 평가기초자료를 국가적 차원에서 또는 내부평가자나 프로그램 관리자가 다양하고 자세히, 그리고 주기적으로 수집하여야 할 것이다.

Campbell[10]은 모든 정책과 프로그램에 대한 의사결정은 지속적인 실험에 기반하여 이루어져야 함을 주장하면서 평가설계의 엄격함을 강조하는 반면, Cronbach[13]는 방법론의 엄격성보다도 평가는 프로그램 후원자 및 이해관계자의 정보요구를 반영해야 한다는 유용성을 강조하면서 목적에 따라 실험설계 및 절차와 방법이 유연하게 변할 수 있음을 주장하고 있다. 또한 Weiss and Bucuvalas[42]는 155명의 의사결정자에 대한 조사를 통해 그들이 보고서를 의사결정과정에서 이용할 때에 결과의 진실성과 효용성을 가장 크게 고려함을 알아내었다. 결국 프로그램평가에서 평

가방법을 선택하고 적용할 때에는 방법론적인 엄격함과 그 결과가 나타내는 유용성을 동시에 고려해야 함을 알 수 있다. 즉 최소한의 과학적 기준을 만족시키면서 동시에 의사결정자에게 최대한 유용한 결과를 낼 수 있는 평가방법의 선택이 필요한 것이다.

### 3.3. 지원시스템

#### 3.3.1 평가정보시스템

앞에서도 말했듯이 프로그램의 영향도를 평가하기 위한 첫번째 전제조건은 프로그램 실행에 대한 자료가 확보되어야 한다는 것이다. 프로그램의 종료시점에서 자료를 수집하는 것의 가장 큰 문제점은 응답자들이 회고적으로 답해야 하므로 기억이 정확하지 못하거나, 응답과정에서 의도적으로 편의를 일으킬 수 있다는 것이다[16]. 따라서 평가결과의 타당성 및 신뢰성이 결여되거나 프로그램의 영향도가 과대 또는 과소 평가되기 쉽다. 이러한 문제점을 해결할 수 있는 한가지 방법은 사례 10에서 나타난 것처럼 실시간(Real Time)평가를 수행하는 것이다. 실시간 평가란 프로그램 진행중에 수행하는 평가를 말하며, 실시간 평가가 사전평가나 사후평가에 대하여 가지는 장점은 첫째, 프로그램의 종료후에는 얻을 수 없는 정확하고, 실제적인 자료와 전문가 및 다른 프로그램 참가자의 현장적 의견수집이 가능하고, 둘째, 평가결과를 프로그램 관리자나 스태프에게 피드백함으로써 프로그램의 성과와 지휘에 영향을 끼칠 수 있으며, 셋째, 평가목표에 따라서 가장 적절한 자료를 가장 적절한 시점에 확보할 수 있다는 것이다[17].

사례 2에서 보면 각 센터마다 내부평가자가 상주하여, 센터의 발전과정과 핵심사건, 센터관계자간 의사소통 유형, 센터의 조직구조와 대학과의

관계, 실현 성과, 그리고 당면 과제 등을 지속적으로 기록하여 다양한 분석방법을 적용가능하게 하고 점진적으로 시행착오를 개선해 나갔으며, 새로운 센터의 기획 및 운영에 큰 도움을 주었다. 사례 1에서는 프로그램 평가를 위한 기초자료나 주기적 자료를 전혀 축적하고 있지 않았기 때문에 촉박한 시간속에서 평가 자료를 수집하는데 큰 어려움을 겪고, 방법론적인 제약에 당면함을 볼 수 있었다. 평가기초자료 수집을 위한 한 가지 방법은 사례 2나 11처럼 내부평가부서를 개별적으로 설립하여 일상적인 관리와 통제, 자문 업무를 수행하고 동시에 외부평가자에게 제공할 평가 관련 정보를 지속적으로 수집하게 하는 것이다. 물론 프로그램 관리자나 실행자가 자료를 기록, 수집할 수 있으나, 표준화되고 체계적인 기록을 위해서는 특별한 교육과 훈련이 필요하고, 또한 편의가 발생가능하므로, 그 자료의 신뢰성도 높지 못하게 된다.

효과적인 평가 자료수집을 위해서 매우 유용하게 이용될 수 있는 것 중의 하나가 경영정보시스템(MIS)이다. MIS는 자료를 체계적으로 조직하여 저장하기 때문에 다양한 정보요구에 따라서 다양한 시점에서 다양한 방법으로 정보를 통합 및 제공 가능하고, 따라서 프로그램 관리자의 의사결정, 보고서 작성, 이해관계자가 요구하는 정보의 제공 및 프로그램의 효과성과 효율성의 평가에 이용 가능하다[32].

### 3.3.2 평가예산 및 기간

평가를 통해서 프로그램의 기획 및 실행을 향상시키기 위해서는 평가가 프로그램의 일부가 되어야 한다[30]. 이미 UNESCO와 ILO에서는 평가가 프로그램의 통합부분이 되어 프로그램의 준비과정에서 평가의 준비를 포함시키고 있다[9]. 프로그램 관리자나 정책결정자는 프로그램 평가

가 정부자원 배분과정에 기여한다는 확신을 가지고 높은 수준의 객관적 평가를 추구해야 하며, 이의 촉진을 위해서 가장 적절한 방법과 유능하고 공정한 평가자의 선택 및 이들에 대한 보상과 인센티브 제공에 유의해야 한다[18].

평가에 드는 비용과 평가결과의 질 사이에는 상관관계가 존재한다. 사례 10에서 보듯이 체계적이고 지속적으로 연구과제의 선정부터 중간 점검을 거쳐 프로그램 종료 후의 영향도까지의 전과정을 모니터링하기 위해서는 짧은 기간 동안 행해지는 사후 평가보다 훨씬 많은 비용이 든다. 또한 사례 1에서 보듯이 실험설계의 정교성과 평가결과의 통계적 유의성을 높이기 위해서는 비용과 시간이 더 많이 든다. 다양한 원천에서 자료를 수집할수록, 다양한 분석방법을 적용할수록, 그리고 전문성과 인지도가 높은 외부평가집단을 초빙할수록 평가비용과 기간이 더 많이 필요하게 된다.

임박한 정치적 의사결정과정에 맞추어 평가결과를 제출하기 위해 급하게 평가를 설계하면, 사례 1에서처럼 더 바람직한 방법이 있음에도 불구하고, 부득이 결과의 신뢰도와 타당도가 낮은 방법에 의존하게 된다. 특히 혁신 및 연구개발 프로그램 특징 중의 하나는 연구개발 성과가 나타나기까지 시간지연이 있다는 것이다[5,14]. 따라서 사례 4에서처럼 너무 일찍 프로그램의 성과를 평가하다보면 단기적 효과만을 측정하여 프로그램의 전체 성과를 과소평가할 수 있고, 분석 내용 및 유의한 방법론의 적용에도 제약을 받게 된다. 결국 효과적인 프로그램 평가를 위해서는 원하는 평가결과의 질적 수준에 비례하여 예산을 할당해야 하며, 다양한 자료 수집과 가장 적절한 분석방법의 적용을 위한 충분한 시간이 주어져야 할 것이다. 의사결정 시점에서부터 일정기간 앞서서 평가결과가 도출되어 그 내용이 정책결정자 및 관련 이해관계자에게 충분히 배포되고 이해되는

것이 의사결정과정에서 평가결과가 활용되기 위한 하나의 필요조건임을 알아야 한다[37].

### 3.4. 평가결과의 활용

평가는 일종의 정치적 활동으로서, 새로운 정책의 개발, 프로그램의 설계와 실행 및 프로그램의 관리와 운영 향상 등에 의도적으로 영향을 끼치려고 하는 목적이 있으며, 이를 통하여 인간생활의 상태를 향상시키려고 한다. 평가의 목적은 이렇게 그 결과가 프로그램 의사 결정에 영향을 끼치도록 하는 것이지만, 아무리 평가목적이 명확하고 평가설계가 우수하다하더라도 평가결과가 활용되지 않는 수도 많다 [28,41]. 사례 2, 6, 8, 11 그리고 12와 같이 평가결과가 프로그램 관리자나 의사결정자에 유용한 정보를 제공하여 프로그램의 관리와 실행을 개선하거나, 사례 4, 7, 그리고 9에서처럼 프로그램의 지속 및 중단 결정, 새로운 방향 설정 등에 지대한 영향력을 행사할 수 있는 반면, 사례 1과 3에서와 같이 평가결과가 전혀 의사결정과정에서 반영되지 않을 수도 있다.

평가결과가 활용되는 유형은 크게 세가지로 나눌 수 있다[20]. 첫째는, 프로그램의 채택이나 지속, 중단 결정 등과 관련하여 평가결과가 직접적으로 의사결정에 영향을 끼치는 것으로, 수단적 활용이라고 하며, 사례 6, 8, 그리고 11 등이 이에 해당한다. 둘째는, 프로그램의 의사결정에 간접적으로 영향을 미치는 것으로, 사람 또는 집단을 특정 사회 문제에 대하여 민감하게 만들고 결과의 축적을 통해 향후 프로그램의 개발 등에 영향을 끼치는 것으로, 개념적 활용이라고 하며, 사례 2, 5, 7, 그리고 9 등이 이에 해당한다. 셋째는 정치적 과정 속에서 자신 또는 집단의 지위나 위상을 방어 또는 공격하기 위해 평가결과를 제시하는 것으로, 설득적 활용이라고 하며, 사례 1과

4가 이에 해당한다.

평가결과와 활용과 관련하여 주의 깊게 고려해야 할 사실 중의 하나는, 평가란 다양한 이해관계자가 관련된 정치적 활동이라는 것이다. 평가에 관심을 갖고 영향을 받는 집단은 다수이고, 대체로 이들은 경쟁적, 상충적 시각을 갖고 있다. 이해관계자로는 정책결정자, 프로그램 및 평가 후원자, 프로그램 목표 집단, 프로그램 관리자 및 스텝, 평가자, 그리고 기타 관련 이해관계자 집단 등이 있으며, 이들 중에서 평가결과와 활용에 주도적인 영향력을 행사하는 집단의 관심, 요구 및 이해관계를 파악하고 평가에 반영하는 것이 필요하다. 사례 1에서는 강력한 영향력이 있는 정책결정자와 프로그램 후원집단 및 경쟁 집단의 반대로 평가결과와는 관계없이 프로그램이 중단되었고, 사례 3에서는 프로그램에 의하여 직접적으로 영향받는 이해관계자 집단의 협상 및 로비 능력에 의하여 평가 결과에 관계없이 프로그램이 지속되었으며, 반면에 사례 7에서는 프로그램 후원 집단의 대표들이 평가과정에 참여함으로써 평가 결과의 수용도를 높였고, 사례 11에서는 평가문제의 결정시 다양한 이해관계자가 참석한 토론회의 개최 및 중요한 집단에 대한 개별 면담을 통하여 이들의 의견을 최대한 반영하고자 노력하였다.

이들 이해집단의 요구를 가장 잘 반영할 수 있는 방법 중의 한 가지는 평가과정에 평가결과와 이용자를 직접 참여시키는 것이다. 참여를 통하여 이들이 중요하게 생각하는 문제에 평가가 지향되었다는 확신을 심어주고, 연구에 대한 관심과 헌신을 확보할 수 있으며, 평가자는 실제적인 평가 문제를 정확히 파악하여 평가의 현실적 유용성을 향상시키고 이해관계자와의 가치 및 시각 차이를 조정할 수 있다[36]. 비록 이들이 직접적으로 참여하지 못할지라도, 평가자는 빈번한 의견 교환을 주선하여, 그들의 관심사를 항상 파악하고 있어야

한다. 평가시기의 선택에 있어서는 사례 6과 7에서와 같이 문제의 사회적 중요도와 의사결정 시점을 인식하여 의사결정자들이 평가결과를 적시에 이용할 수 있도록 평가결과의 제출시기와 정치적 의사결정 시점을 맞추어야 한다.

평가의 정치적 특성을 강조하는 또 한가지 사실은 평가결과가 통계적으로 유의하고, 일반화 가능성이 높아도 정책적 의미가 없을 수 있다는 것이다[35]. 즉 평가과정에서 정책결정자의 관심과 요구를 반영하지 못했거나, 결과의 해석, 배포 및 전달이 적절히 이루어지지 않은 경우에 그와 같이 될 수 있다. 평가자가 중요하게 생각하는 것과 정책결정자 또는 프로그램 관리자가 알고 싶어하는 것은 다를 수 있고, 수치자료와 전문 용어로 가득 찬 두꺼운 보고서에 아무도 읽으려 하지 않을 것이며, 명확한 추천대안이나 행동지침이 없을 경우 의사결정자들은 평가결과를 어떻게 활용해야 할지 모를 것이다. 따라서 사례 6과 7에서처럼 보고서에 대한 발표회나 토론회를 거쳐 정책적 의의를 찾아보고, 그 결과를 학술지 등에 배포하거나, 사례 9에서와 같이 최종 보고서가 제출되기 전에 보고서 초안에 대하여 그 결과를 이용할 집단 및 영향받는 집단과 사전 검토 및 의견 교환 과정을 갖는 등의 추가적인 노력이 필요하다. 결국 평가결과의 활용을 위해서는 단순한 보고서의 제출로는 부족하고, 이해관계자의 요구에 맞도록 비전문용어로 간결 명료하게 요약하여 다양한 방법으로 전달하는 세심한 주의와 노력이 필요하다.

Coleman[11]도 평가결과의 활용을 저해하는 요인들은 평가의 기획과 의사결정을 위한 정책적 의의를 도출하는 과정에 밀집해 있다고 지적하면서, 평가문제의 구체화와 평가방법의 결정 과정에서 다양한 이해관계자의 의견과 관심을 반영하고, 평가결과와 추천안의 제안시 의사결정자의 영향력에 의하여 변화가능한 문제와 조건에 집중할

것을 권장하고 있다. 평가결과는 자동적으로 의사결정에 영향을 끼칠 것이고, 연구의 우수성이 결과의 활용을 보장하리라는 단순한 생각은 버려야 한다[39]. 평가결과가 의사결정에 직접적으로 이용되기 위해서는 많은 의도적이고 추가적인 노력이 평가자보다도 평가후원자 및 프로그램 관리자와 스텝 수준에서 더 필요하다. 주요 이해관계자가 필요로 하는 평가문제와 의사결정 시점의 명확한 인식, 적절한 연구의 설계와 실행, 명확한 결과와 적시의 발표, 평가자와 이해관계자간 가치의 일치, 결과와 특정 평가상황과의 부합, 의사결정자에 대한 외부압력의 부재, 연구를 위한 충분한 자원과 권한 등 복잡하고 다양한 요인들이 평가결과의 활용에 영향을 끼치므로, 평가가 반드시 정책을 결정해야 한다는 생각은 바람직하지 못하고, 그 보다는 평가결과는 본질적으로 정치적 의사결정과정에서 하나의 투입요소라는 생각을 가지고, 평가자는 최고의 정보와 지식을 제공하려고 노력해야 할 것이다[32,41].

### 3.5. 평가에 대한 인식

평가결과의 질적 수준은 평가방법 못지않게 평가자료의 정확성과 신뢰성에 의존한다. 평가에 이용되는 자료는 정성적 자료와 정량적 자료가 있으며, 서로의 장단점이 있으므로 함께 이용해야 한다. 아무리 정량적인 자료라 할지라도 그 이면에는 주관적인 측면이 반영되어 있기 마련이고, 정량적 자료의 적용에는 한계가 있다[26]. 결국 정성적 자료와 정량적 자료의 질과 정확성은 모두 응답자의 태도나 자세에 의하여 영향을 받게 되어 있음을 알 수 있다. 사례 2, 5, 그리고 6에서 보듯이 프로그램의 평가가 자신들의 활동을 정당화하고 지속적인 지원을 얻는데 도움이 되거나, 프로그램의 관리와 운영상의 문제점을 해결하여

프로그램의 효과성 제고에 기여하거나, 또는 프로그램의 존립여부에 위협을 주는 것이 아닌 앞으로의 방향설정이나 전략수립에 도움이 될 때, 피평가 집단은 평가에 대하여 저항하거나 기만적인 답변을 하지 않고 유용한 정보를 신속하게 제공하는 등 평가에 대하여 저항하거나 기만적인 답변을 하지 않고 유용한 정보를 신속하게 제공하는 등 평가에 대하여 우호적 태도를 갖게 되는 것을 알 수 있다.

의사결정자도 평가에 대하여 관심을 가지고 공정하고 객관적인 평가가 될 수 있도록 충분한 자원 및 인센티브를 제공하고, 평가의 기획, 실행 및 결과 도출과정에 적극적으로 참여하여 평가를 지원해 줄때, 보다 더 유용한 결과가 산출되고, 따라서 결과의 활용도도 제고될 것이다. 평가결과의 활용이 평가결과의 신뢰도와 유용성에 의하여 영향 받는다는 사실을 고려해볼때, 평가결과의 활용과 평가에 대한 인식은 상호작용하여 서로를 강화시킬 수 있음을 알 수 있다.

이러한 평가문화의 정착을 위해서 가장 필요한 것 중의 한가지는 평가의 제도화이다. 사례에서 볼 수 있듯이 많은 국가에서는 일시적이건, 지속적이건 의사결정 및 프로그램의 관리 및 운영 향상을 위하여 평가를 수행하고 있으며, 특히 사례 1, 2, 3, 7, 10, 그리고 11에서는 프로그램의 평가를 공식적으로 규정하고 있다. 스웨덴의 경우 프로그램의 평가가 하나의 전통으로 자리잡고 있어서, 프로그램의 정치적 중요성에 대한 치열한 논쟁 속에서도, 평가가 아무런 저항이나 무리없이 공정하게 진행되고 있음을 볼 수 있다. 사례 10에서는 프로그램의 시작부터 지속적으로 평가를 실시하여, 좀더 실제적이고 정확한 자료를 이용한 평가를 하고 있으며, 그 결과는 정책결정자 뿐만 아니라 프로그램 관리자 및 스텝에게도 도움을 주고 있다. 이렇게 평가를 제도화하여 평가문화를 정착시킴으로써, 이

해관계자 모두가 평가에 대한 바른 인식을 갖도록 하고, 프로그램 관리자 및 스텝의 전문성 향상도 도모할 수 있으며, 프로그램의 시작과 동시에 평가를 준비하여 프로그램 실행동안 평가 기반 자료를 충실히 수집함으로써 신뢰할 수 있고 유용한 평가 결과를 산출할 수 있고, 궁극적으로는 평가결과의 활용을 촉진할 수 있다.

## 4. 외국사례의 교훈 및 평가지침 개발

지금까지 외국의 혁신 및 연구개발 프로그램 평가 사례를 소개하고 사회과학 분야에서 발전된 평가연구이론을 기반으로 평가목적, 기본시스템, 지원시스템, 평가결과의 활용 그리고 평가에 대한 인식으로 구성되어 있는 프로그램 평가틀을 이용하여 사례들을 분석해 보았다. 여기에서는 이 분석 결과를 토대로 향후 혁신 및 연구개발 프로그램의 평가시 유용하게 도움을 줄 수 있는 교훈 및 지침을 제시해 보고자 한다.

### 4.1. 평가목적

① 프로그램 평가의 기획단계에서는 우선적으로 프로그램 목표와 활동의 명확화가 필요하다. 이를 위해서는 평가성조사가 유용하게 이용될 수 있다.

② 평가 목적을 구체화해야 하며, 이 과정에는 정책결정자, 프로그램 관리자 및 스텝, 프로그램에 의하여 영향받는 집단 등 다양한 이해관계자의 참여가 필요하고, 이들이 원하는 정보와 평가자가 중요하게 생각하는 정보 사이의 조정과 합의 과정이 필요하다.

③ 평가목적과 단계에 따라서 적합한 평가방법 및 자료의 선택이 필요하다. 즉 프로그램 모니터링을 위해서는 정성적 평가 방법이 적합하고, 프

로그램 영향도 및 효율성 평가를 위해서는 정성적 평가 방법과 정량적 평가 방법을 모두 이용되 후자에 더 큰 비중을 두는 것이 바람직하다. 또한 정확한 영향도 평가를 위해서는 모니터링 정보가 선행적으로 확보되어야 한다.

#### 4.2. 기본시스템

① 평가결과의 객관성과 공정성 제고를 위해서는 외부평가가 바람직하며, 내부 평가부서를 설치하여 프로그램 관련 정보 및 평가 자료 등을 외부평가자에게 지원하도록 하는 것이 좋다. 외부평가자의 선정시에는 평가자의 전문성과 인지도를 중요하게 고려해야 하며, 평가 문제에 따라서 적합한 다른 평가자를 위촉하는 것도 좋은 방법이다. 부득이 내부평가를 실시해야 할 때는 합의된 원칙이나 규칙을 따르도록 하며 평가 절차를 공개하는 것이 바람직하다.

② 자체평가는 내재적인 장단점을 잘 비교하여 수행여부를 결정해야 하며, 의사결정을 위한 주요 기준으로 이용하기 보다는 다른 평가에 추가하여 보조적인 자료로 이용하는 것이 바람직하다.

③ 정성적인 평가 방법과 정량적인 평가 방법은 서로가 가지지 못하는 장점이 있다. 평가방법의 선택시에는 평가문제, 이해관계자의 정보요구, 확보가능한 자료의 종류 등을 고려하여 가장 적합한 방법을 선택해야 한다.

④ 평가시에는 가능한 한 다양한 자료원천과 평가방법을 이용하고, 이들을 서로 비교하여 평가결과를 제시함으로써, 결과의 타당성을 향상시키고 여러가지 측정 오류를 제거할 수 있다.

⑤ 프로그램 효과의 정확한 판단을 위해서는 정교한 실험설계 방법을 이용하는 것이 바람직하고, 이의 설계시에는 평가에 주어진 예산, 시간, 확보가능한 자료 및 이해관계 집단의 협조 정도

등을 고려해야 한다. 특히 프로그램의 커버리지가 80% 이상일 때에는 재귀적 통계에 의한 평가설계만이 가능하지만, 그렇지 않은 경우에는 비교집단을 이용한 평가설계가 더 바람직하다.

⑥ 평가방법의 선택시에는 방법론적인 엄격함과 결과의 유용성이 상충적인 경우가 많으므로, 최소한의 과학적 기준은 만족시키면서 정책결정자에게 최대한 유용한 결과를 줄 수 있는 방안을 강구해야 한다.

#### 4.3. 지원시스템

① 정확하고 유용한 평가결과의 획득을 위해서는 프로그램의 시작부터 지속적으로 평가를 위한 기초 자료를 수집하는 것이 필요하다. 이것은 내부 평가부서나 교육과 훈련을 받은 프로그램 집행자가 할 수 있고, 경영정보시스템(MIS)을 이용하면 자료수집의 효율성 향상외에도 다양한 장점이 있다.

② 실시간(Real Time) 평가는 수집된 자료의 질과 프로그램 관리 및 운영 향상 측면에서 사전평가나 사후평가보다도 많은 이점을 가지고 있으므로 예산이 허용한다면 권장할 만하다.

③ 평가가 프로그램의 설계 및 실행에 기여하고 궁극적으로 의사결정에 유용한 정보를 제공하기 위해서는 프로그램 활동 중의 일부로 통합되어야 한다. 이를 위해서는 평가에 대한 정책결정자의 지원과 헌신, 그리고 독립된 예산의 배정이 필요하다.

④ 평가결과의 질과 평가에 드는 비용사이에는 상관관계가 존재한다. 좀더 신뢰할 수 있고 가치가 큰 결과를 위해서는, 정교한 평가설계, 다양한 자료원천의 활용 및 적절한 분석 방법의 적용이 필요하고, 따라서 비용이 커지게 된다.

⑤ 평가시기 및 기간을 결정할 때에는 연구개



발의 특징 중의 하나인 시간지연성과 정치적 의사결정 시점을 동시에 고려하여야 한다. 너무 일찍 또는 단기간내에 평가가 수행되면 얻을 수 있는 정보가 제한되고, 너무 늦게 또는 장기간동안 평가가 수행되면 의사결정시에 활용되지 못한다.

#### 4.4. 평가결과의 활용

① 평가는 다양한 이해관계자 집단이 존재하는 정치적 활동임을 인식하고 평가의 기획 실행 및 결과 도출과정에서 이들의 시각과 가치를 반영해야 한다. 항상 그들의 요구를 이해하고 고려하며, 되도록 평가 과정에 핵심적인 이해관계자 집단을 참가시킴으로써 평가결과의 효용성을 향상시키고 수용도를 높일 수 있다.

② 평가결과는 문제의 사회적 중요도와 의사결정 시점을 고려하여 의사결정자들이 충분히 이해하고 이용할 수 있도록 적시에 제출되어야 한다.

③ 평가결과의 제출시에는 단순한 사실의 나열이 아니고, 의사결정자가 필요로 하는 정보에 대한 정책적 의의를 찾아 이해하기 쉽도록 간결하면서도 명료하게 전달해야 한다.

④ 정책적 대안을 제시할 때에는, 이해관계자 집단의 요구를 반영해야 할 뿐만 아니라, 정책 결정자가 영향력을 행사하여 변화시킬 수 있는 부분에 집중해야 한다.

⑤ 평가결과가 자연적으로 의사결정과정에 영향을 끼치기를 기대해서는 안된다. 평가자는 평가결과가 본질적으로 정치적 의사결정 과정에 투입되는 하나의 요소일 뿐임을 인식하고 신뢰도와 효용성이 높은 최고의 정보를 제공하려고 노력해야 하며, 평가결과의 직접적인 활용을 위해서는 평가자 뿐만 아니라 평가 후원자 및 프로그램 관리자와 스텝의 적극적이고 의도적인 노력이 필요하다.

#### 4.5. 평가에 대한 인식

① 평가대상집단은 평가가 프로그램의 정당성 제시 및 효과성 제고에 기여한다고 생각할 때 평가에 대하여 긍정적으로 반응하며 양질의 정확한 자료를 제공한다.

② 의사결정자가 평가에 대하여 관심을 가지고 충분한 예산을 배정하고, 우수한 평가에 대한 보상과 인센티브를 제공하며, 평가과정에 적극적으로 참가할수록, 보다 유용한 결과가 산출되고, 따라서 평가결과의 활용도도 향상된다.

③ 평가대상집단 및 의사결정자의 평가에 대한 인식을 변화시키기 위해서는 평가문화의 정착이 필요하고, 이를 위한 방법 중의 하나가 평가의 제도화이다. 평가를 프로그램 활동 중의 일부분으로 규정함으로써, 프로그램 관리자 및 스텝의 전문성 향상, 양질의 평가기반자료 수집, 그리고 평가에 대한 긍정적 태도 확립 등이 가능하게 해야 한다.

## 5. 결 론

본 고에서는 미국, 독일, 프랑스, 스웨덴, 오스트레일리아, 영국, EC, 그리고 일본에서 수행되었던 12의 국책 혁신 및 연구개발 프로그램의 평가사례를 소개하고, 이를 사회과학분야에서 발전되어온 평가연구이론을 기반으로 하여 종합적으로 분석하였다. 이 과정에서 프로그램 평가의 주요 구성요소 및 이들간의 관계를 나타낸 평가를 위한 개념적 틀에 따라 분석 및 설명을 체계적으로 하려고 노력하였다. 또한 사례들의 분석 결과를 토대로 향후 우리 나라에서 연구개발 프로그램의 평가를 할때 유용하게 이용할 수 있는 평가지침을 제시하였다.

오늘날 세계각국은 기술경쟁이 가속화됨에 따라 국가의 경쟁력 제고와 경제적 목표 달성을 위

해서는 산업에서의 새로운 기술개발과 혁신이 필수적임을 공통적으로 인식하고 있다. 동시에 투입 가능한 자원이 제한된 반면, 연구개발 비용이 급증하고, 성과가 나타나기까지의 시간이 장기화되다 보니, 정부활동의 정당성 제시와 효율성 제고 요구가 이와 함께 증가하고 있다. 따라서, 정부정책의 책임성을 증대하고, 자원배분의 효율성을 높이며, 연구자원 집중투입을 위한 우선순위 분야 결정을 위하여 선진국들은 연구개발평가를 일회성 활동이 아닌 하나의 제도로서 정착시키려고 노력하고 있다.

이제 정부의 연구개발정책은 더이상 신념과 근거없는 기대에 의한 것이 아닌, 실증적 자료에 기반한 “평가에 의한 관리(Management by Evaluation)”가 되어야 한다. 연구개발 프로그램의 평가를 위해서는 사회과학 분야에서 발전된 다양한 평가방법의 이용이 가능하다. 평가연구는 주로 사회학, 정치학, 경제학, 심리학, 행정학, 교육학, 사회복지학, 보건학, 사회개발학 등의 분야에서 1970년대와 1980년대에 걸쳐 급속히 발전했으며, 연구개발 분야에서도 정책결정을 위한 유용한 정보제공수단이 될 수 있다. 우수한 평가전문가의 확보를 위해서는, 국내 각종 공식 및 비공식 교육훈련기관에 연구개발평가 과정을 설치 및 확대하고, 국비유학생의 선발이나 박사후 과정 해외연수 등에 연구개발 평가전문가 양성을 포함시키며, 평가담당 기관 및 평가자의 교육훈련을 강화하는 등의 적극적인 노력이 필요하다[6].

연구개발 활동은 투입에서 성과발현까지 시간적 지연이 있고, 가시적인 성과 뿐만 아니라 측정하기 어려운 파급효과도 있다. 따라서 단기적이고, 쉽게 측정할 수 있는 효과만으로 프로그램의 성과를 판단하는 것은 지양해야 하며, 일차적인 목표뿐만 아니라 사회경제적 영향도까지 포괄하여 평가할 수 있는 평가방법 및 지표의 개발에

노력을 기울여야 한다.

평가는 본질적으로 정치적 활동이어서 의사결정을 지원하고, 자원의 재분배를 유도하며, 자원 사용에 대한 정당성을 제시하기 위해서 수행된다. 양질의 평가결과를 확보하기 위해서는, 프로그램 관리자 및 스텝, 정책결정자, 프로그램 수혜집단 등 다양한 이해관계자 집단의 적극적인 참여와, 평가의 목적이 사후적인 결과의 통제가 아닌 프로그램의 효과성과 효율성 제고에 있다는 건설적인 인식의 확산이 필요하다. 또한 정책결정자는 평가에 필요한 예산과 인력을 지원하고, 우수한 평가에 대한 보상과 인센티브를 제공하며, 정책결정시 평가결과에 기반함으로써 의사결정의 합리성을 높이고, 프로그램 이해관계자집단에게 평가의 필요성 및 유용성을 인식시켜야 한다.

## 參 考 文 獻

1. 김계수, 김재영, 이민형, “국책연구개발사업의 평가시스템에 관한 연구” 「과학기술정책」, 4권, 2호(1992), pp.65-92
2. 김명수, 「공공정책평가론」, 박영사, 1987.
3. 노화준, 「정책평가론」, 법문사, 1986.
4. 윤문섭, “연구평가의 동향과 평가 시스템의 구축 방안” 「과학기술정책동향」, 1권, 15호(1991), pp.3-6.
5. 이진주, “연구개발계획과 연구관리시스템” 「기술관리」, 2권, 5호(1984), pp.48-53.
6. 정보통신연구관리단, 「정보통신 연구개발의 종합평가시스템 발전방안 워크샵」, 1993.
7. Anderson, J., *New Approaches to Evaluation in UK Research Funding Agencies*, London, U.K., 1989.
8. Barbarie, A., *Evaluation of Programmes Promoting Technological Innovation in*

- Canada, prepared for the workshop on Evaluation of Government Innovation Programmes, OECD, Paris, 1987.
9. Bobe, B. and H. Viala, *One Decade of R&D Evaluation at the European Commission(1980-1990)*, pp.1990.
  10. Campbell, D. T., "Reforms as Experiments," *American Psychologist*, Vol.24(1969), pp. 409-429.
  11. Coleman, J. S., *Policy Research in the Social Science*, General Learning Press, Morristown, NJ, 1972.
  12. Cook, T. D., "Dilemmas in Evaluation of Social Programs." in M. B. Brewer and B. E. Collins(eds), *Scientific Inquiry and the Social Science: A Volume in Honor of D. T. Campbell*, Jossey-Bass., San Francisco, 1981, pp.257-286.
  13. Cronbach, L. J., *Designing Evaluation of Educational and Social Programs*, Jossey-Bass, San Francisco, 1982.
  14. Gee, R. E., "The Opportunity Criterion-A New Approach to the Evaluation of R&D" *Research Management*, Vol.15, No.3 (1972), pp.64-71.
  15. Gibbons, M., "The Evaluation of Government Policies for Innovation" in J. D. Roessner(ed), *Government Innovation Policy: Design, Implementation, Evaluation*, St. Martins Press, Inc., New York, 1988, pp.135-146.
  16. Gibbons, M. and L. Georghiou, *Evaluation of Research: Synthesis Report*, Organization for Economic Cooperation and Development, Directorate for Science, Technology and Industry, OECD, 1986.
  17. Hobday, M., "Evaluating Collaborative R&D Programmes in Information Technology: The Case of the U. K. Alvey Programme" *Technovation*, Vol.8(1988), pp. 271-298.
  18. Kostoff, R. N., M. A. Averch, and D. E. Chubin, "Research Impact Assessment: Introduction and Overview," *Evaluation Review*, Vol.18, No.1(1994),pp.3-10.
  19. Krull, W., D. Sensi and O. Sotiriou, *Evaluation of Research & Development: Current Practice and Guideline*, Commission of European Communities, 1991.
  20. Levidon, L. C. and E. F. X. Hughes, "Research on the Utilization of Evaluations: A Review and Synthesis," *Evaluation Review*, Vol.5(1981). pp.525-548.
  21. Mark, M. M. and T. D. Cook, "Design of Randomized Experiments and Quasi-Experiments" in L. Rutman(ed.), *Evaluation Research Methods: A Basic Guide(2nd ed.)*, Sage Publications, Inc., Newbury Park, California, 1990, pp.65-120.
  22. McKeon, R. and J. A. Ryan, "Evaluation of Programs Technological Innovation-The Australian Experience," *Research Policy*, Vol.18(1989), pp.379-388.
  23. Meyer-Krahmer, F., "Evaluation of Industrial Innovation Policy: Concepts, Methods and Lessons," in J. D. Roessner (ed.), *Government Innovation Policy: Design, Implementation, Evaluation*, St. Martins Press, Inc., New York, 1988, pp. 121-133.

24. Meyer-Krahmer, F. and P. Montigny, "Evaluations of Innovation Programmes in Selected European Countries," *Research Policy, Vol.18(1989)*, pp.313-332.
25. OECD, *Evaluation of Programmes Promoting Technological Innovation*, General Distribution, Organization for Economic Cooperation and Development, Paris, 1989.
26. Ormala, E., "Nordic Experiences of the Evaluation of Technical Research and Development," *Research Policy, Vol.18(1989)*, pp.333-342.
27. Patton, M. Q., "Data Collection Options, Strategies and Cautions," in L. Rutman (ed.), *Evaluation Research Methods: A Basic Guide(2nd ed.)*, Sage Publications, Inc., Newbury Prk, California, 1990, pp. 39-64.
28. Patton, M. Q., *Utilization-Focused Evaluation(2nd ed.)*, Sage Publication, Inc., Newbury Park, California, 1986.
29. Raizen, S. A. and P. H. Rossi(eds.), *Program Evaluation in Education: When? How? To What Ends?*, National Academy Press, Washington, DC, 1981.
30. Roessner, J. D., "Evaluation of Government Innovation Programs : Introduction," *Research Policy, Vol.18(1989)*, pp. 309-312.
31. Roessner, J. D., "Evaluatin Government Innovation Programs : Lessons from the S. Experience," *Research Policy, Vol.18 (1989)*, pp.343-359.
32. Rossi, P. H. and H. E. Freeman, *Evaluation: A Systematic Approach(4th ed.)*, Sage Publications, Inc., Newbury Park, California, 1991.
33. Rutman, L., "Evaluability Assessment," in L. Rutman(ed.), *Evaluation Research Methods: A Basic Guide(2nd ed.)*, Sage Publications, Inc., Newbury Park, California, 1990, pp.27-38.
34. Scrivin, M., "The Methodology of Evaluation," in C. H. Weiss(ed.), *Evaluating Action Programs: Readings in Social Action and Education*, Allyn and Bacon, Inc., Bosto, 1972, pp.123-136.
35. Sechrest, L. and W. H. Yeaton, "Magnitudes of Experimental Effects in Social Science Research," *Evaluation Review, Vol. 6(1982)*, pp.579-600.
36. Smith, N. L. and D. N. Caulley(eds.), *The Interaction of Evaluation and Policy: Case Reports from State Education Agencies*, Northwest Regional Educational Laboratory, Portland, OR, 1982.
37. Solomon, M. A. and S. M. Shortell, "Designing Health Policy Research for Utilization," *Heath Policy Quarterly, Vol.1(1981)*, pp.237-261.
38. Tanaka, M., "Japanese-Style Evaluation Systems for R&D Projects : The MITI Experience," *Research Policy, Vol.18 (1989)*, pp. 361-378.
39. Useem, M. and P. DiMaggio, "An Example of Evaluation Research as a Cottage Industry : The Technical Qalty and Impact of Arts Audience Studies," *Sociological Methods and Research, Vol.7 (1978)*, pp.55-84.
40. Weiss, C. H., *Evaluation Research Methods of Assessing Program Effective-*

- ness, Prentice-Hall, Inc., New Jersey, Englewood Cliffs, 1972.
41. Weiss, C. H., "Increasing the Likelihood of Influencing Decisions," in L. Rutman(ed.), *Evaluation Research Methods: A Basic Guide(2nd ed.)*, Sage Publications, Inc., Newbury Park, California, 1990, pp. 159-190.
42. Weiss, C. H. and M. J. Bucuvalas, "Truth Tests and Utility Tests : Decision-Makers' Frames of Reference for Social Science Research," *American Sociological Review*, Vol. 45(1980), pp.302-313.
43. Williams, W. and R. F. Elmore(eds.), *Social Program Implementation*, Academic Press, New York, 1976.