

## 개발평가의 품질보증(Quality assurance) 방안

조 정 명 (KOICA 평가실 과장)

### 목 차

1. 품질보증 논의 배경
2. 품질보증을 위한 노력
  - 2.1 OECD/DAC 평가품질 기준
  - 2.2 KOICA의 개발평가 품질관리 기준
  - 2.3 경제학적 개념 활용
  - 2.4 시민사회사업 사례
3. 향후과제 및 연구제언

## 1. 품질보증 논의 배경

개발원조사업 평가의 주요 목적은 크게 두 가지로 정리해 볼 수 있다. 첫째는 책임성 accountability 확보의 목적이며, 둘째는 학습 learning의 목적이다. 개발원조를 시행하는 기관은 평가를 제도화함으로써 임무수행의 성과를 납세자 tax payer 또는 자금 지원자 fund provider에 게 보고할 의무를 충족시킬 수 있다. 또한, 평가를 통해 개발사업을 개선할 수 있는 방법 및 지식을 생산하고, 생산된 지식을 향후 전략 및 사업계획 수립에 반영할 수 있다. 평가가 목적에 충실하고 그 결과가 향후 의사결정에 제대로 활용되기 위해서는 평가 과정 및 결과에 대한 신뢰성과 정확성이 보장되어야 한다. 따라서 평가의 품질 보증은 원조의 책임성 강화를 위한 증거기반

마련에 도움이 되며, 향후 사업추진을 위한 유용한 참고자료 생성에 기여하는 것이다.

최근 국제사회의 개발환경 변화에 따라 평가의 역할과 중요성이 증대되고 있다. 개발원조 파트너들은 원조의 효과성에 대한 파리선언Paris Declaration on Aid Effectiveness(2005), 아크라 행동강령Accra Agenda for Action(2008) 등을 통해 개발 목표와 의무를 공유하고, 새로운 원조방식에 대한 합의Consensus를 이루고 있다. 이러한 환경의 변화에 따라 성과중심관리, 상호책임성 등이 더욱 강조되고 있으며, 이에 따라 평가의 역할도 증대되고 있다. 평가는 정책 결정자에게 의사결정의 기반이 되는 신뢰할 만한 정보를 제공하고, 협력 파트너들이 개발성과에 대한 상호 책임성을 증명하는데 매우 중요한 역할을 하기 때문이다. 개발평가의 수행방식 또한 이러한 변화의 흐름을 반영해야하며, 평가의 역할과 중요성이 강조됨에 따라 평가의 품질보증에 대한 요구도 높아지고 있다.

이에 본 글은 지난 6월 ‘개발평가의 품질관리’를 주제로 개최된 2012년도 2차 개발평가 세미나<sup>1)</sup> 결과를 토대로, 평가 품질보증을 위한 다양한 주체들의 노력을 정리하고, 향후 개발평가 품질관리 추진 방안에 대해 논하고자 한다.

## 2. 품질 보증을 위한 노력

### 2.1. OECD/DAC 평가품질 기준

평가의 품질 보증을 위한 노력은 다양하게 시도되어 왔다. 먼저 국제적인 측면에서는 OECD/DAC개발원조위원회 Evalnet평가네트워크의 개발평가의 품질기준Quality standards for development evaluation 수립을 들 수 있다. Evalnet은 개발평가의 품질을 높이고, 궁극적으로는

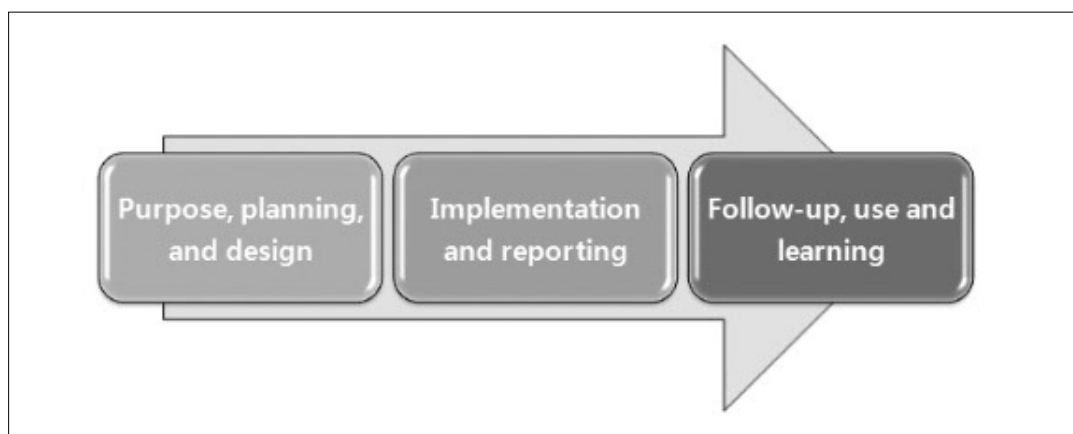
---

1) KOICA 평가실은 2011년부터 개발평가 관련 지식공유를 위해 분기별 개발평가 세미나를 개최하고 있다. 2012년도에는 ①OECD/DAC 평가네트워크(Evalnet)이슈 공유, ②개발평가 품질관리(Quality Assurance)방안, ③평가결과 활용도(Feedback)제고 방안, ④수원국 개발평가 역량강화 방안 등을 주제로 세미나를 개최할 계획이다. 지난 6월에 개최된 2차 세미나에서는 KOICA의 평가품질관리 기준을 공유하고, 컬럼비아 대학교 SIPA(School of International and Public Affairs) Cristian Pop-Eleches 교수, 국제빈곤퇴치기여금 지원 사업인 프로젝트 말라위의 김현철 책임자 등이 발표 연사로 참가하여 학계와 시민사회단체의 평가의 품질관리를 위한 노력을 공유하였다.

질 좋은 평가가 성과 제고에 기여하는 것을 목표로 개발평가 품질 기준을 수립하였다. 이 기준은 개발환경 변화에 따른 새로운 평가방식에 대한 요구를 반영하여 수립되었으며, 국제사회에서 공통으로 적용할 수 있는 ‘양질의 평가Quality evaluation’에 대한 가이드라인을 제시한다는 데 의의가 있다. 3년간의 조사 및 협력 파트너들 간의 협의를 통해 2010년 완성되었으며, 의무 적용 사항은 아니나 국제사회에서 개발협력 평가의 품질 보증을 위한 지침으로 인식되고 있다. Evalnet은 특히 이를 평가의 품질 점검, 바람직한 평가에 대한 정보제공, 평가역량강화 교육, 관련 정책 수립 등에 참고로 활용할 것을 권장하고 있다. 다만, 국가나 지역적 특색을 고려하여 적절하게 적용되어야 하며, 평가의 목적에 따라 다양하게 활용될 수 있다.

동 기준은 DAC에서 제시하는 개발평가의 원칙DAC Principles for the Evaluation of Development Assistance(1991) ; 공정성Impartiality, 독립성Independence, 신뢰성Credibility, 유용성Usefulness 등을 기반으로 수립되었다. 또한 일반적인 평가과정에 따라 평가의 목적정의Defining purpose, 계획Planning, 설계Designing, 집행Implementation, 보고Reporting 및 학습Learning 등의 순으로 구성되어 있다. 평가단계에 따라 반드시 고려해야 하는 요인을 정리하고, 각 요인별로 충족되어야 하는 기본사항을 정리하여 평가자 또는 평가관리자가 품질 좋은 평가 생산을 위한 일종의 체크리스트로 활용할 수 있다. 단, Evalnet은 동 기준을 평가수행을 위한 기본지침이나 가이드라인으로 활용하는 것은 적절하지 않다고 권고한다.

〈그림 1〉 일반적 평가수행 단계



DAC의 평가 품질기준은 평가의 가장 처음 단계인 목적, 계획, 설계 단계에서 고려되어야 할 사항을 다음과 같이 정리하였다. 먼저, 평가의 배경과 목적이 구체적으로 명시되어야 한다. 평가가 어떠한 목적으로, 언제, 누구를 위해 시행되며, 어떻게 사용될 것인지를 분명히 해야 한다.

다음으로는 평가의 대상과 범위를 분명히 해야 한다. 평가 대상에는 사업의 개입 논리 및 배경 이론 등에 대한 설명도 포함되어야 하며, 평가의 범위에는 평가대상이 되는 기간, 사용 예산, 지역, 집단, 조직구성, 정책적·제도적 환경 등이 고려되어야 한다. 평가성Evaluability 또한 사전에 점검 되어야 한다. 평가성이란 평가가 가능한지 여부를 살펴보는 것으로 평가대상이 명확한지, 사업의 결과가 입증될 수 있는지, 평가가 사업의 이해관계자나 정책결정자들의 의문 사항에 답할 수 있는 최선의 방법인지 등을 점검하는 것이다. 평가성이 낮다고 판단되는 평가는 이 단계에서 중지하는 것이 바람직하다. 평가성에 문제가 없다고 판단되면, 평가가 답하고자 하는 이슈와 질문을 고려하여 관련 이해관계자가 모두 포함되었는지 점검한다. 새로운 개발협력 환경에서 강조되는 원조조화, 수원국 정책과의 일치 측면에서 다른 파트너와의 협력이 필요한지 확인하여, 필요시 공동평가를 실시한다.

평가 목적과 계획이 완료되면, 평가 질문개발, 평가기준 적용 등의 평가 설계 단계에 착수한다. 평가 질문은 평가를 통해 달성하고자 하는 목적과 답변하고자 하는 핵심사항을 질문의 형태로 반영하는 것이라고 생각하면 쉽다. 평가 기준은 DAC의 기준인 적절성Relevance, 효율성Efficiency, 효과성Effectiveness, 영향력Impact, 지속가능성Sustainability이 기본원칙이되 평가의 목적에 따라 융통성 있게 적용하도록 한다. 평가방법은 평가의 목적, 범위, 질문에 따라 적합한 방식을 결정한다. 평가 방법을 기술할 때에는 평가 설계 및 자료수집 방법에 대한 정당성을 포함해야하며, 성과 지표는 SMART Specific, measurable, attainable, realistic, timely한 것을 활용하도록 한다. 또한 자료 수집 시에는 성별, 소득별로 분리하여 남, 녀간 또는 소득계층간의 사업성과 차이를 평가 결과에 포함하도록 하는 것이 좋다.

평가의 목적, 계획, 설계가 마무리 되면 행정적인 사항을 고려해야 한다. 가령, 평가에 필요한 물적, 인적 자원이 충분한지, 평가 관리체계는 투명하고 신뢰할 만한지 점검한다. 최종적으로는 위에서 검토된 모든 사항을 반영하여 'Terms of Reference(TOR)'로 문서화 하도록 한다. TOR은 'Scope of work'이나 'evaluation mandate'라고 불리기도 한다.

다음으로는 실질적인 평가 이행과 보고의 단계 이다. 이행단계에서는 가장 먼저 투명한 조달 절차를 통해 평가팀을 선발하도록 한다. 평가팀은 평가전문가와 관련된 기술 전문가가 조합된 형태가 바람직하다. 평가 구성원 간의 성별균형과 수원국의 전문가를 평가팀에 포함하는 것이 좋다. 평가는 사업의 계획, 집행으로부터 독립적으로 이루어져야 한다. 다시 말해, 사업과 관련된 정책수립, 운영 및 관리에 포함되었던 관계자들은 평가팀에 포함될 수 없으며, 평가팀은 사업 관계자들의 간섭 및 통제로부터 자유로워야 한다. 또한 사업 관련 자료에 대한 자유로운 접근도 허용되어야 한다.

평가 이행단계에서 평가팀은 사업과 관련된 모든 이해관계자와의 인터뷰를 통해 그들의 의견을 수렴하도록 한다. 물론 이해관계자 선정 기준도 명확히 제시되어야 한다. 의견수렴 과정에서 평가에 참여한 사람들에 대한 권리를 보장토록 하며, 필요시 인터뷰 대상자들에 대한 익명성과 비밀유지를 보장한다. 평가는 약속된 시간과 예산 안에서 수행되도록 하며, 환경과 조건에 변화가 발생할 시에는 평가자는 반드시 평가 의뢰자에게 보고하도록 하며, 필요시 관련사항을 이해관계자들 간 협의를 통해 조정한다.

평가결과의 보고 시에는 다음 사항을 반드시 포함하도록 한다 : 보고서 요약, 개발협력 사업 추진 환경(정책적·제도적·사회문화적 환경, 이해관계자의 참여 등), 사업개입 논리, 정보출처, 평가 방법론, 분석결과, 평가의 한계 등 분석결과를 제시할 때 주의해야 할 점은 평가를 통해 확인된 사실Fact과 고찰 및 제언사항Recommendation을 반드시 명확하게 분리하여 제시해야 한다는 것이다. 또한 결과보고서 작성 시에는 당초 TOR상에서 약속했던 바, 즉 평가를 통해 달성하고자 했던 목적과 답변하고자 했던 의문사항들이 충분히 충족되도록 한다. 결과에 제시한 판단이나 제언사항에 평가팀 구성원 간의 의견 불일치가 있을 경우, 이를 사실대로 언급하도록 한다. 마지막으로 평가와 관련된 이해관계자들은 평가 결과보고서 초안에 대해 의견을 제시할 수 있으며, 최종 보고서는 이해관계자들의 의견을 수렴, 반영하여 완성한다.

평가의 마지막 단계는 후속조치와 학습의 단계이다. 평가가 약속된 시간에 완성, 배포되어 이해관계자들이 평가의 결과와 제언사항을 유용하게 활용할 수 있도록 한다. 또한 평가를 통해 제안된 사항에 대해서는 체계적으로 후속조치가 이루어 져야 한다. 이는 평가결과의 환류제도 Feedback 등 공식적인 대응장치 및 이행과정에 대한 모니터링 등을 필요로 한다. 이러한 측면에서 효율적인 평가 결과의 확산방법 또한 고려되어야 한다. 내·외부적으로 평가결과에 대한 접근성이 보장되어야 하며, 학습효과의 극대화를 위해서는 평가와 관련된 직접적인 이해관계자 뿐만 아니라 다른 국제개발협력 파트너들과도 평가 결과를 공유하는 것이 바람직하다.

## 2.2. KOICA의 개발평가 품질관리 기준

개발평가의 품질보증을 위한 기관차원의 노력으로는 KOICA의 개발 평가의 품질 관리 기준을 들 수 있다. KOICA는 OECD/DAC Evalnet의 품질기준을 준용하여 2011년 국내 원조기관 중 최초로 기관차원의 개발평가 품질관리 기준을 수립하였다. 품질관리 기준을 평가의 절차에 따라 평가 설계의 타당성Relevance of evaluation design, 평가수행의 일관성 및 투명성Consistency and transparency of evaluation implementation, 평가결과의 신뢰성Credibility of evaluation results 등

으로 구분하고, 각 기준에 따라 세부 관리항목과 세부 점검요소를 설정하였다. 동 기준은 현재 평가실에서 실시하는 모든 평가의 전 과정에서 품질 관리를 위한 지침으로 활용되고 있으며, 최종 보고서 심사 시 평가 질Quality 부분 심사기준으로도 엄격히 적용되고 있다.

평가 설계의 타당성과 관련해서는 세부 관리 항목을 다음과 같이 네 가지로 정리하였다 : 이해관계자 파악, 대상 분석, 목적 및 범위설정, 방법론 제시. 최근 평가 결과의 신뢰성 및 독립성 보장을 위해 외부 독립기관에 의한 평가가 장려되고 있으며, KOICA 평가실의 평가 또한 특수한 경우를 제외하고 외부평가로 시행되고 있다. 외부기관에 의한 평가의 경우, 독립성 측면에서는 높이 평가 되나, 평가대상 및 제반환경에 대한 이해 부족으로 유용한 평가 결과 도출이나 학습효과 측면에서는 그 효과가 제한적 일 수 있다. 이러한 한계를 극복하기 위해서는 평가 착수 이전에 평가 의뢰자-수행자 간의 충분한 의사소통을 통해 평가 설계에 대한 합의를 이루는 것이 중요하다. 체계적이고 구체적인 평가 설계는 품질 높은 평가 결과 도출 및 평가의 활용성 제고에 기반이 되기 때문이다. 이러한 이유로 KOICA 평가실에서는 TOR 작성, 착수보고회 등의 절차를 통해 외부 평가팀과 충분히 의견을 교류함으로써 평가의 목적에 타당한 평가 설계를 도출하고자 노력하고 있다. 평가 설계의 타당성 검증을 위한 세부 관리 항목 별 점검 요소는 <표 1>와 같다.

**< 표 1> KOICA 평가 품질관리 기준 : 평가 설계의 타당성**

세부관리 항목	세부점검 요소
평가이해관계자 파악	• 이해관계자의 범위, 종류, 중요도 등을 파악했는지 점검한다.
	• 위 사항을 고려하여 이해관계자 분석을 실시하였는지 점검한다.
평가대상 분석	• 사업의 법적 근거, 상위계획여부, 수행 체계 등을 파악했는지 점검한다.
	• 사업의 목적, 기존의 성과모형 및 성과지표를 파악했는지, 이를 분석하여 평가기준에 반영하였는지 점검한다.
	• 기존의 목적 및 성과지표가 적절하지 않다면 이를 평가팀이 구성하여 제시 하고 있는지를 점검한다.
	• 재구성과정에서 관련자와 협의하는 과정을 거쳤는지 점검한다.
	• 유사사례 평가에 대한 조사를 수행하였는지 점검한다.
평가의 목적 및 범위 설정	• 의뢰자의 평가의 목적과 수요를 명확히 파악하고 있는지 점검한다.
	• 공식적으로 합의된 평가목적을 계획서에 명시하고 있는지 점검한다.
	• 평가의 요소 및 기준을 평가팀과 의뢰팀간 합의과정을 거쳤는지 점검한다.
	• 평가제약요소가 충분히 파악되고 이를 계획에 반영하여 수립하였는지 점검한다.

세부관리 항목	세부점검 요소
평가 방법론 제시	• 평가모형의 종류 및 방법을 제시하고 있는지 점검한다.
	• 평가 매트릭스를 주요 질문을 포함하여 설계하였는지 점검한다.
	• 제시된 평가매트릭스가 평가수행에 바로 적용할 수 있도록 충분히 구체적인지 점검한다.
	• 제시된 평가모형 및 방법이 시간상, 예산상의 제약을 고려하여 구성되었는지 파악한다.
	• 앞서 “이해관계자 파악 및 분석”을 바탕으로 주요 정보원(Key informants) 및 기타 이해관계자를 확인하고, 샘플링을 수행하였는지를 점검한다.
	• Sampling의 표본추출, 구성, 수혜자를 포함한 다양한 이해관계자가 평가에 참여하도록 계획 되었는지 점검한다.

평가수행의 일관성 및 투명성과 관련한 세부관리 항목은 네 가지로 정리하였다 : 국내 문헌 및 자료 조사, 현지조사 계획 및 수행, 조사내용 분석 수행, 평가자의 태도 및 윤리준수. 평가의 수행방식은 평가대상의 특성이나 목적에 따라 다양할 수 있다. 그러나 평가의 품질보증을 위해서는 평가수행과 관련해 개인적인 업무방식 차이로 인해 발생하는 업무혼선을 최소화하는 것이 바람직하다. 이에 평가실은 일관적인 평가수행 가이드라인 제공을 위해 2012년 ‘평가수행 길라잡이’를 발간하였다. 또한, 평가수행 단계의 품질관리와 진행사항에 대한 정보공유를 위해 중간 보고회 등을 개최한다. 평가수행의 일관성 및 투명성 항목의 세부점검 요소를 정리하면 아래와 같다.

〈표 2〉 KOICA 평가 품질관리 기준 : 평가수행의 일관성 및 투명성

세부관리 항목	세부점검 요소
국내 문헌 및 자료 조사	• 평가대상 사업 분야의 자료 리스트를 파악하여 제시하고 있는지를 점검한다.
	• 평가에 필요한 추가자료 목록을 제시하고 있는지를 점검한다.
	• 자료의 출처들을 기록하고 있는지를 점검한다.
현지조사 계획 및 수행	• 현지조사를 통해 확인하고자 하는 항목을 제시하고 있는지 점검한다.
	• 앞서 ‘국내문헌 및 자료조사’ 시 평가에 필요한 추가자료 목록에 포함되는 항목인지 점검한다.
	• 현지에서 획득할 자료의 수집 상세계획(기간, 대상, 자료 종류 등)을 제시하고 있는지 점검한다.

세부관리 항목	세부점검 요소
현지조사 계획 및 수행	• 앞서 ‘평가방법론’의 Sampling 구성과 일치하는지를 점검한다.
	• 상기 현지조사방법이 예산, 시간 등을 고려하여 현실적으로 제시하고 있는지 점검한다.
	• 현지조사 수행을 위한 설문 면담자 훈련, 시범설문조사 등이 사전에 이루어지고 준비하고 있는가를 점검한다.
	• 현지조사 항목이 미확인된 경우 대안을 제시하고 있는지 점검한다.
조사내용 분석 수행	• 앞서 설계단계에서 제시한 ‘평가방법론’ 과 일치하는지 점검한다.
	• 차이가 있다면 그 이유를 제시하고 있는지 점검한다.
평가자의 태도 및 윤리준수	• 피평가자의 의견을 존중하고 경청하는지를 점검한다.
	• 수원국의 문화적, 성인지적, 환경적 차이 <sup>2)</sup> 를 인지하고 이를 준수하는지 점검한다.
	• 사전 평가자 행동강령(Code of Conduct) <sup>3)</sup> 을 숙지하고 있는지 점검한다.

마지막으로 결과보고 단계의 평가결과의 신뢰성 관련 기준은 다음의 세 가지로 정리할 수 있다 : 평가결과의 객관성 제시, 평가 제언의 유용성 검토, 보고서 작성의 성실성, 평가결과 보고는 평가의 완성도 결정에 크게 영향을 미친다. DAC의 품질기준에서도 강조된 바와 같이 평가 분석결과를 객관적으로 전달하는데 있어 관찰 사실과 평가자의 고찰 및 의견을 구분하는 것이 매우 중요하다. 또한 독립 평가자의 경우 평가결과의 유용성 제고를 위해서는 사업 집행의 제도적 환경과 자원의 한계 등을 고려하여 분석결과 및 제언사항을 도출하는 것이 필요하다. KOICA에서는 최종보고회 등을 통해 이해관계자의 의견을 최종적으로 수렴하고, 이를 반영하여 유용한 평가결과를 도출할 수 있도록 노력하고 있다. 이 부분에서 한 가지 강조하고자 하는 점은 최종 보고 단계에서의 이해관계자 의견수렴은 평가결과의 객관성 및 독립성 문제와는 다른 차원에서 이해되어야 한다는 것이다. 이는 분석결과에 대한 개입이라기보다는 궁극적으로 평가결과를 사용하게 될 이해관계자의 요구를 반영하여, 평가 결과의 쓰임을 제고하기 위함이다. 아울러 평가 설계상에 제시되었던 평가의 목적이 평가 결과 분석에서 모두 충족되었는지 확인하여 최종보고서의 완성도를 높이하고자 하는 것이다. 평가결과의 신뢰성 판단을 위해 점검해야 할 요소를 정리하면 아래 <표3>과 같다.

2) “성인지 가이드라인(성인지 담당관)” 및 “지속가능한 개발원조를 위한 KOICA 환경가이드라인(2008.12)” 참조

3) DAC 의 평가품질관리기준 핸드북의 행동강령(Evaluation Ethics) 활용



〈표 3〉 KOICA 평가 품질관리 기준 : 평가결과의 신뢰성

세부관리 항목	세부점검 요소
평가결과의 객관성 제시	• 평가결과가 앞서 ‘평가방법론’ 및 ‘조사내용 분석수행’에 제시한 분석 방법을 통해 도출되었는지 점검한다.
	• 평가방법론이 변경되었다면, 그 이유를 제시하고 있는지 점검한다.
	• 각 평가결과별 근거자료를 제시하고 있는지 점검한다.
평가제언의 유용성 검토	• 평가 주요발견사항(findings), 제언사항(Suggestion), 기타의견 <sup>4)</sup> 등 을 구 별하여 제시하는지를 점검한다.
	• 각 제언사항이 해당 이해관계자별로 제시되었는지를 점검한다.
	• 제언사항이 공식적으로 합의된 평가 목적과 일치하는가를 점검한다.
	• 제언사항은 현실적이며 의뢰기관이 수행가능한지를 점검한다.
보고서 작성의 성실성	• 평가의 제약요건을 점검하여 제시하고 있는지를 점검한다.
	• 의뢰자의 보고서 수정사항 요청에 대한 대응을 객관적으로 제시하고 있는지를 점검한다.
	• 보고서 제출기한, 수정안 보고 등을 적시에 하는지를 점검한다.
	• 평가보고서의 자료 및 인용 출처를 명확히 하고 있는지를 점검한다.
	• 평가보고서(국·영문)의 용어 <sup>5)</sup> 및 어휘사용이 의뢰기관의 분야에 맞추어 적절하게 사용되었는지 점검한다.

위와 같은 KOICA 평가품질 관리 기준은 앞에서 언급한 바와 같이 최종 보고서 심사단계 뿐만 아니라, 평가의 전 과정에 걸쳐 평가의 품질보증을 위한 기본 지침으로 활용된다. 개발평가의 품질보증 노력을 제도화 하기 위해 현재 KOICA 평가실에서 수행하는 모든 평가는 계약 체결시 품질관리 관련 사항으로 동 기준을 포함하도록 하고 있다. 이밖에도 지속적인 대내외 교육을 통해 개발평가의 품질보증의 필요성 및 방법을 확산할 계획이다.

4) 평가 주요결과는 평가분석 및 방법론을 통해 도출된 객관적인 결과이며, 제언사항은 평가결과를 토대로 평가자의 가치판단, 주관적인 의견이 개입된 바람직한 방향제시를 의미한다.

5) KOICA번역본 “평가 및 성과중심관리 주요 용어사전(DAC)” 참조

## 2.3. 경제학적 개념 활용

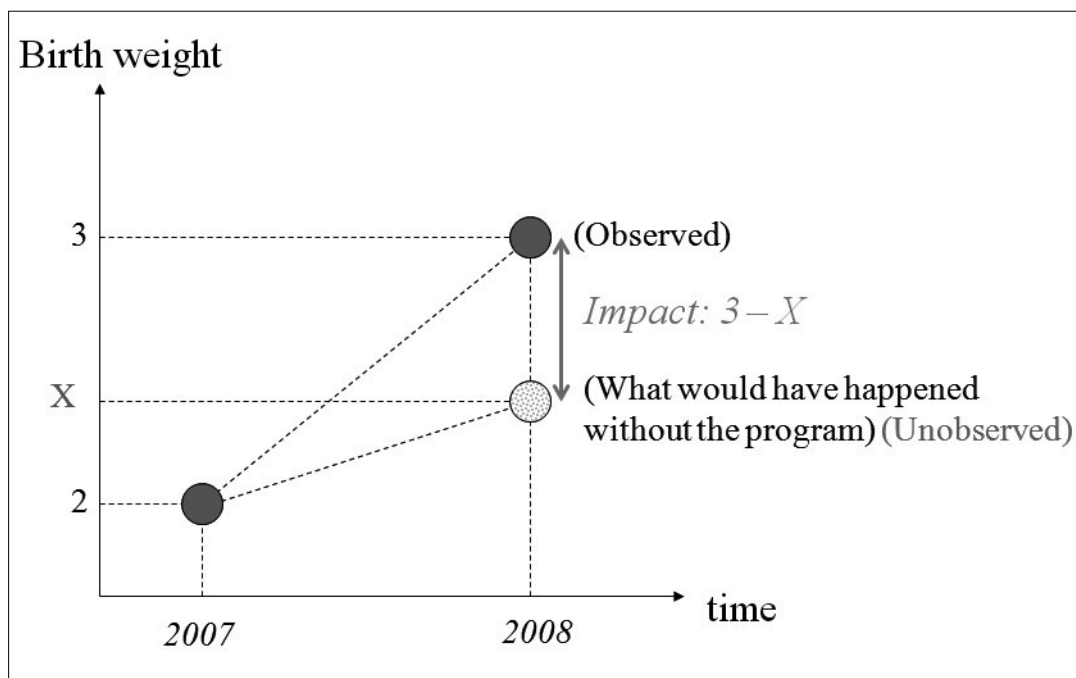
최근 평가를 통한 성과입증 및 성과의 계량화에 대한 대내외적 요구가 높아짐에 따라 경제학적 개념 및 도구를 활용한 개발평가가 많이 시도되고 있다. 가장 대표적인 사례 중 하나가 영향평가Impact evaluation 이다. 영향평가는 과학적인 설계를 바탕으로 하는 고차원적인 평가 방법 중 하나로, 특정 성과지표에 대해 사업의 수혜를 받은 집단과 사업의 수혜를 받지 않은 집단, 즉 대조군Counterfactual의 사업 후 결과를 비교하여 사업의 순효과Net effect를 추정하는 방법이다.

컬럼비아 대학교 SIPA(School of International and Public Affairs) Cristian Pop-Eleches 교수는 제 2차 개발평가 세미나에서 영향평가를 양질의 평가결과를 도출할 수 있는 대표적인 방법 중 하나로 소개하고, 실제 사업의 예를 활용하여 영향평가의 개념과 장점을 설명하였다.

인도네시아 보건사업 예를 통해 영향평가의 개념을 정리해 보도록 하겠다. 인도네시아 Mentawai 섬에 1/4의 아이들이 12세가 되기 전에 사망하며, 주요 사망 원인 중 하나가 말라리아라는 문제의식에서 말라리아 퇴치 사업을 기획하였다. 말라리아 감염의 취약집단은 산모와 유아Infant로 구분할 수 있는데, 산모의 말라리아 감염은 저체중 유아를 출산하여 신생아 및 영아사망률에 큰 영향을 미치며, 유아의 말라리아 감염은 생존에 치명적이다. 이에 5세 이하 아동 및 임신여성을 대상으로 말라리아 모기장 배포 사업을 실시하였으며, 이 사업의 성과를 측정하기 위해 영향평가를 시도하였다.

기본적인 영향평가 방법은 사업 착수 이전에 사업 수혜집단과 인구·사회문화적 특성이 유사한 비교집단, 즉 대조군을 설정하고, 사업수혜집단과 대조군에서 사업의 성과지표(인도네시아 사업의 경우 출산체중)에 대한 기초선 조사Baseline survey와 종료선 조사End-line survey를 실시한다. 사업 종료 후에 수혜집단과 대조군의 출산체중 변화를 비교하여 사업의 순효과를 추정하는 것이다. 이와 같은 개념을 도식화하여 나타내면 아래 그림과 같다.

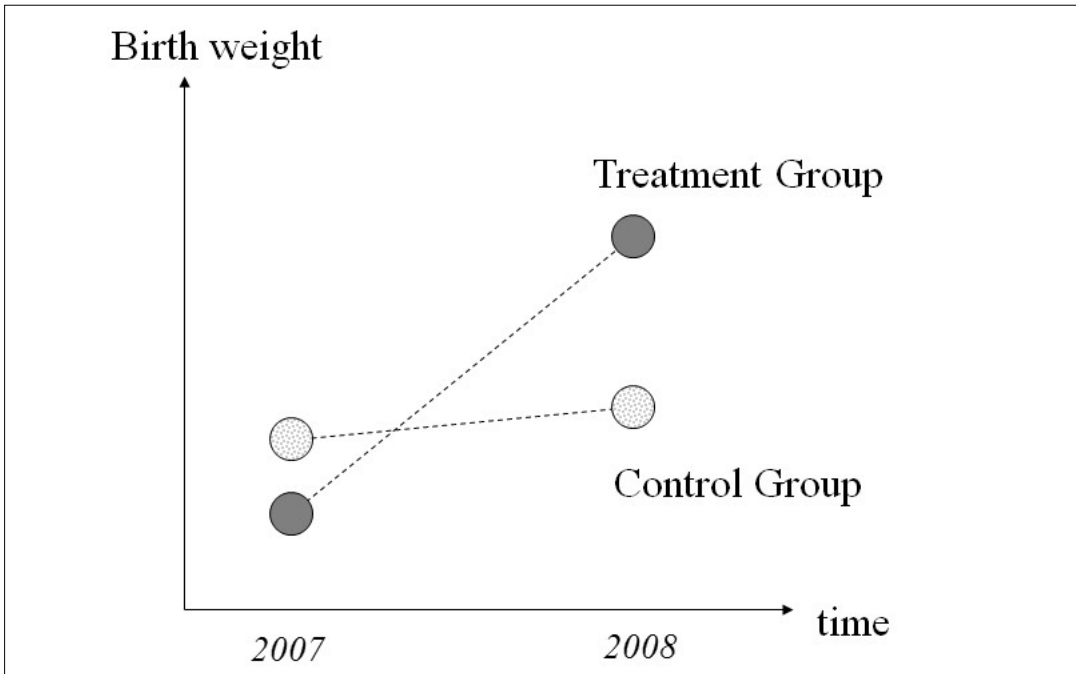
〈그림 2〉 영향평가의 개념



출처 : Cristian Pop-Eleches 교수 발표자료

그러나 비실험적Non-experimental 환경에서 인간과 지역사회를 대상으로 하는 협력 사업의 특성상 위와 같은 결론 도출은 비현실적일 수 있다. 또한 사업의 수혜여부를 제외하고 유사한 인구·사회적 특성을 가진 대조군을 발견해 내는 것은 거의 불가능하다. 가령, 본래 출산체중이 높은 인구집단을 대조군으로 설정하여 수혜집단과 비교했을 때에는 표본의 편중Selection bias으로 인한 왜곡된 결과가 도출될 수 있다. 따라서 영향평가의 이론에 근거하여 가장 사실 값에 가까운 순효과 측정을 위해서는 현실적인 영향평가 방법 고안 및 대조군 설정이 매우 중요하다.

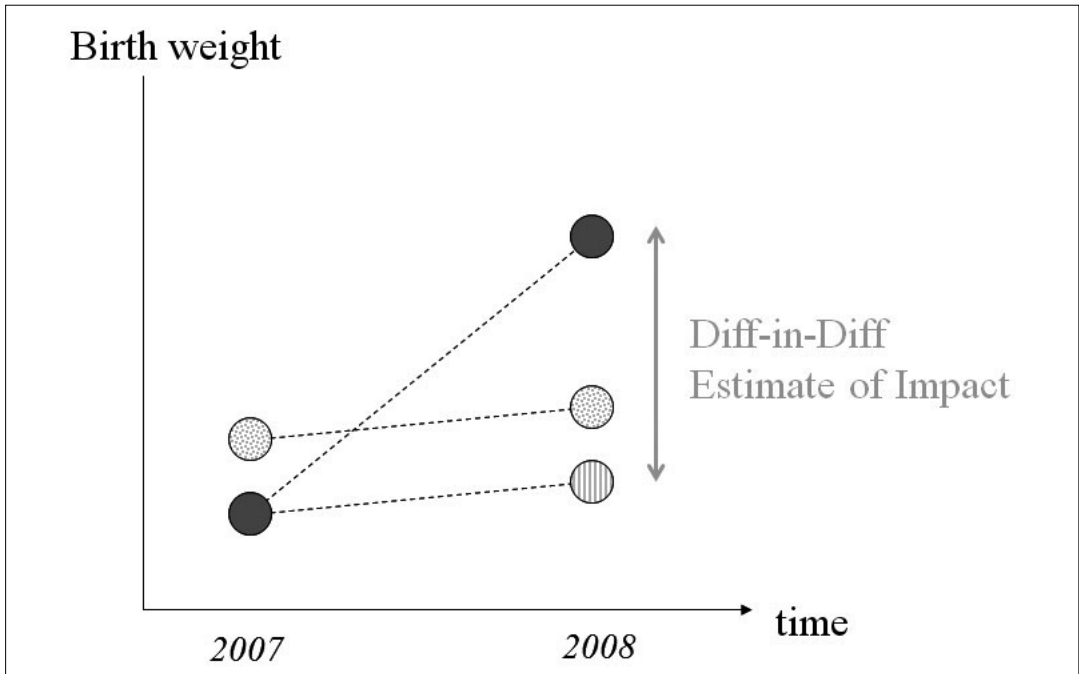
〈그림 3〉 표본의 편중



출처 : Cristian Pop-Eleches 교수 발표자료

이와 같은 한계를 극복하기 위해 사용할 수 있는 방법 중 하나가 ‘Difference-in-difference 방법’이다. 일정 기간 동안 사업 이외의 영역에서 수혜집단과 대조군은 유사한 변화를 보였다는 것을 가정하고, 해당기간 동안 대조군의 변화 양상을 관찰하여 이를 수혜집단에 적용 한 후, 수혜집단의 사업 전후 변화의 차이를 통해 사업의 순효과를 추정하는 것이다. 다시 말해, 수혜집단과 대조군에서 성과지표에 대해 사업수행 전 기초선 조사를 실시하고, 마찬가지로 사업 종료 후 두 집단에서 성과지표에 대한 종료선 조사를 실시한다. 두 집단에서 측정된 변화간의 차이를 계산하여 사업의 순효과를 추정하는 방법이다. 이러한 개념을 도식화하여 나타내면 아래 그림과 같다.

〈그림 4〉 Difference-in-difference의 개념



출처 : Cristian Pop-Eleches 교수 발표자료

〈 표 4 〉 Difference-in-difference를 활용한 순효과 추정

Group	Pre-Program birth weight	Post-Program birth weight	Pre-Post Change
1. Treatment	2.0	2.8	0.8
2. Control	2.6	2.7	0.1
3. Estimated Impact			$0.8 - 0.1 = 0.7$

출처 : Cristian Pop-Eleches 교수 발표자료

그러나 이러한 방법도 표본의 편중으로 인해 발생할 수 있는 오류를 완벽히 통제하지는 못한다. 가령, 인도네시아 정부가 사업 대상지의 빈민 여성을 대상으로 무료 산전 진찰 서비스를 제공한다면, 사업의 효과가 과장되게 나타날 것이다. 이러한 문제를 극복하기 위한 또 다른 방법은 '매칭(Matching)'이다. 이는 수혜집단을 유사한 특성을 가진 비수혜 집단, 즉 대조군과 비교하는 방법이다.

〈 표 5 〉 매칭 방법을 활용한 사업성과 추정

Islands Sampled for Program			
Population	Infant mortality rate in 2000	Infant mortality rate in 2005	Free Pigs
2,345	12%	10%	Yes
1,012	15%	14%	Yes
3,890	10%	5%	No
455	8%	8%	No
622	6%	8%	Yes
1,469	10%	10%	No
2,986	12%	12%	No
980	12%	11%	Yes
2,560	9%	8%	No
570	12%	8%	No
820	8%	6%	Yes
1,070	12%	9%	Yes

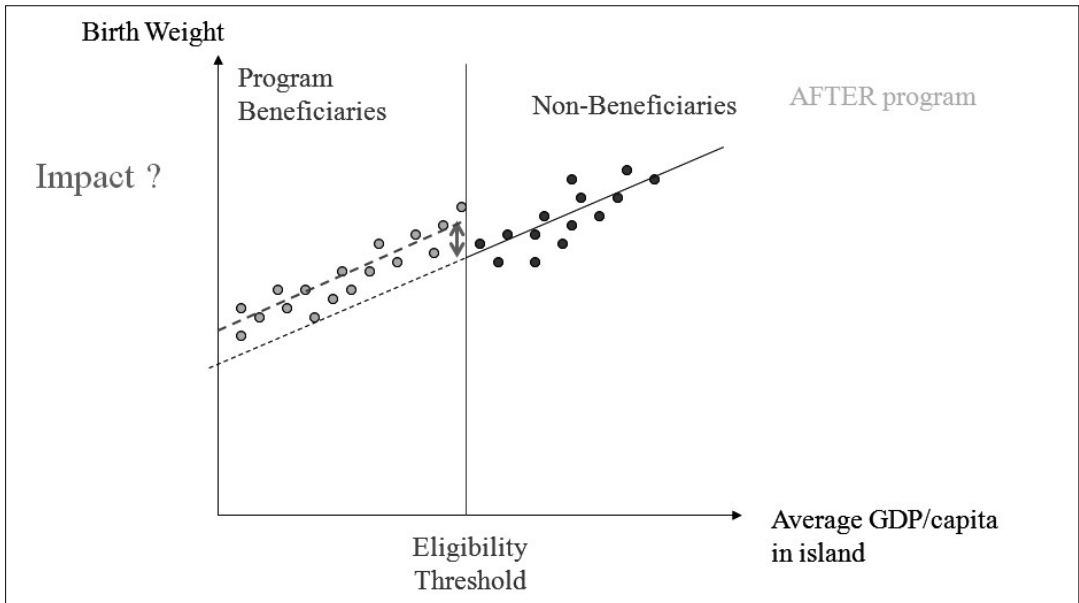
Islands outside of Program			
Population	Infant mortality rate in 2000	Infant mortality rate in 2005	Free Pigs
6,012	3%	2%	No
1,200	14%	13%	Yes
609	6%	8%	No
1,469	10%	10%	No
4,690	7%	5%	Yes
2,525	12%	10%	Yes
1,010	12%	11%	Yes
620	12%	8%	No
3,040	11%	11%	No
590	6%	8%	Yes
2,132	10%	9%	No
4,829	5%	4%	No
5,429	2%	3%	No
2,560	6%	5%	Yes
532	7%	6%	Yes
654	12%	8%	Yes
735	5%	6%	No

출처 : Cristian Pop-Eleches 교수 발표자료

그러나 매칭 방법은 관찰되는 사항에 대해서는 수혜-비수혜 집단 간 비교가 가능하나, 관찰되지 않는 지역사회 지도자의 리더십, 지역 주민들 간의 유대 등과 같이 관찰되지 않는 사항에 대한 통제 및 비교가 불가능하다는 단점이 있다.

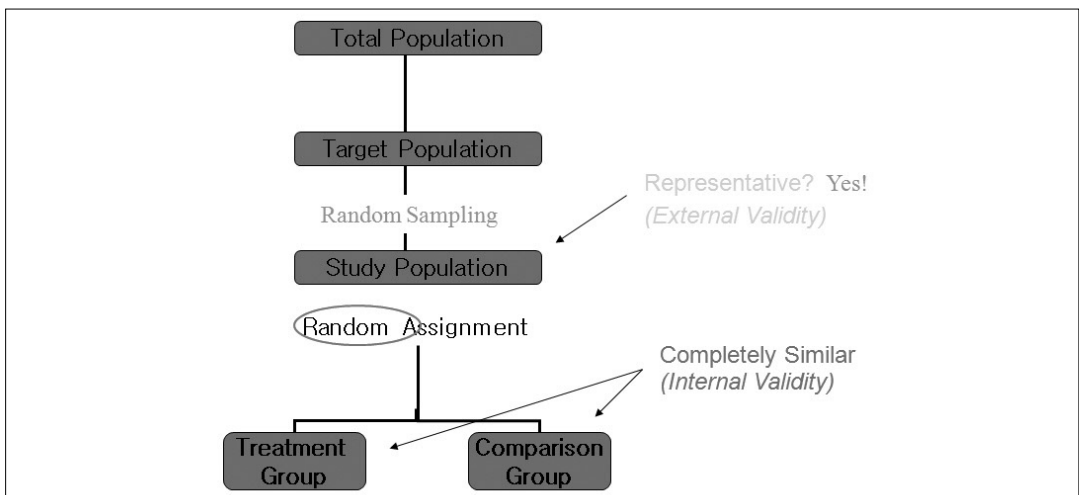
또 다른 유사 집단 간 비교 방법으로는 ‘비연속성 디자인(Discontinuity design)’이 있다. 사업특성상 수혜대상 선정기준이 있는 경우, 가령 연간 소득수준이 200불 미만인 가정을 사업 수혜대상으로 한다거나, 12세 미만 아동을 사업 수혜대상으로 하는 경우에는 비연속성 디자인을 활용할 수 있다. 이는 사업수혜 기준을 중심으로 사업 종료 후 수혜-비수혜 집단 간 성과지표의 차이 비교를 통해 사업의 순효과를 추정하는 방법이다. 비연속성 디자인의 개념을 도식화 하면 아래 그림과 같다.

〈 그림 5 〉 비연속성 디자인을 활용한 사업성과 추정



그러나 지금까지 설명한 방법들은 일정수준에서 표본 편중으로 인한 오류를 극복하는데 한계가 있다. 이러한 오류를 최소화 할 수 있는 방법이 ‘무작위 배정 Random assignment’을 통한 수혜집단 및 대조군 선정 방법이다. 충분히 큰 집단을 무작위로 수혜집단과 대조군으로 구분함으로써, 두 집단이 사업수혜 여부를 제외하고 다른 사항은 평균적으로 공통적인 특성을 띠도록 하는 것이다. 이 경우 사업종료 후 수혜-비수혜(대조군) 집단 간 사업 성과지표 변화의 차이를 비교하면 사업의 순효과를 추정할 수 있다.

〈 그림 6 〉 무작위 배정을 통한 사업성과 추정



사실상 엄밀한 의미의 영향평가는 사업 기획 단계에서부터 고려되어 수행되어야 한다. 그러나 필요에 따라 사업종료 후 후향적인Retrospective 방법으로 영향평가를 수행하기도 하는데, 2011년부터 KOICA에서는 2010년 종료된 에티오피아 가족계획 사업의 사업성과 측정을 위해 후향적 방식의 영향평가를 시범적으로 실시하고 있다. 이 밖에도 2011년부터 성과관리 시범으로 추진 중인 가나 불타지역 식수위생 개선사업을 영향평가가 가능토록 설계하여 추진 중이다. 동 사업은 현재는 사업 착수 단계이나 2013년 이후에 일차적인 평가결과를 도출할 수 있을 것으로 생각된다. 또한 KOICA가 지원하는 시민사회단체 사업 중에서도 아프리카 미래재단의 ‘말라위 치무투, 치투쿠라지역 HIV/AIDS 및 모자보건사업’ 등 영향평가 설계를 도입하여 정확한 사업 성과측정을 시도한 사업도 있다.

이러한 과학적인 평가 설계는 일반 평가 설계보다 더 많은 시간과 예산, 인력 투입이 필요하나, 동시에 정확한 사업성과 측정을 가능케 할 뿐만 아니라 신뢰할 만한 평가기반 자료를 생성하여 결과적으로는 양질의 평가를 도출할 수 있게 한다. 따라서 향후 사업추진을 위한 유용한 참고자료 축적과 전반적인 평가 수준 향상을 통한 품질보증을 위해서는 향후에도 사업 모듈의 검증 등이 필요할 경우에는 영향평가를 시범적으로 적용할 필요가 있다.

## 2.4. 시민사회사업 사례

앞에서 언급한 바와 같이 개발평가의 품질보증은 최종 보고서 심사단계에 이루어지는 것이 아니라, 평가수행과정 전반에 걸쳐 고려되어야 한다. 다시 말해 양질의 평가는 체계적인 평가계획과 지속적인 모니터링이 동반되어야 한다. 사업의 성과를 직접적으로 증명할 수 있는 성과지표 관련 데이터는 신뢰할만하고 유용한 평가결과를 생산하는데 필수요소이기 때문이다. 특히 최근 개발평가에 있어 사업성과 정량화에 대한 대내외적 요구가 증가하고 있어, 사업수행 과정에서도 설문조사 등을 통해 계량적인 자료 수집 필요성이 강조되고 있다. 이에 KOICA 기관 차원에서는 성과관리 시범사업을 통해 사업 초기단계부터 사업 평가 및 모니터링 계획을 구체화하여 성과지표 및 입증수단을 명확히 하고, 지속적으로 관련 데이터를 축적하여 사업성과의 정량화를 강화할 계획이다.

이러한 정확하고 신뢰할만한 자료수집 노력은 시민사회단체에서도 지속되어 왔다. 대표적인 사례 중 하나로 항공연대기금(일명, 빈곤퇴치기여금)으로 지원하는 아프리카 미래재단의 ‘말라위 치무투, 치투쿠라지역 HIV/AIDS 및 모자보건사업’을 들 수 있다. 동 사업의 총 책임자인 김현철 PMProject manager는 제 2차 개발평가세미나에서 신뢰할만한 평가결과 도출의 기반이 되



는 양질의 성과지표 수집을 위한 현장조사(Field survey) 노하우를 공유하였다.

동 말라위 사업은 남학생의 포경수술, 교육, 장학생 제도 등을 통해 사업 대상지역의 HIV/AIDS의 유병률 및 모성 사망 및 영유아 사망률 감소를 목표로 하는 사업으로 2010년부터 현재까지 추진 중이다. 동 사업은 사업성과의 증명을 위해 기획 단계에서부터 영향평가를 고려한 사업 계획을 수립하였으며, 사업 착수 전 기초조사(Baseline survey) 및 지속적인 모니터링을 실시하였다. 특히 정확한 사업성과 관련 데이터 수집을 위해 사업성과 지표를 반영한 설문지 개발에 특별한 노력을 기울였으며, 현장조사로 직접 수집한 자료의 질 관리를 실시하였다. 현장조사는 설문주제의 특성에 따라 단체 설문 및 개인 설문으로 구분하여, 조사방법의 적절성 및 효율성을 기했다.

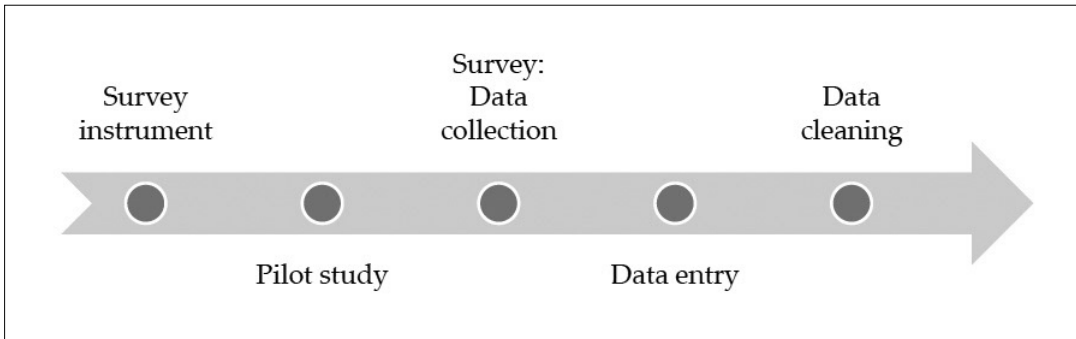
〈 표 6 〉 설문수단

Group survey	Private survey
1. Demographics	7. Selling Condoms
2. Household Assets	8. Time Preference
3. Education	9. Risk Preference
4. General Health	10. HIV/AIDS Knowledge
5. Condoms and HIV/AIDS-related knowledge and attitude	11. Circumcision, Pregnancy and Sexual Practices
6. Participation in Labor	12. Friend Network
	13. Sexually Transmitted Infections and HIV test

출처 : 김현철 PM 발표자료

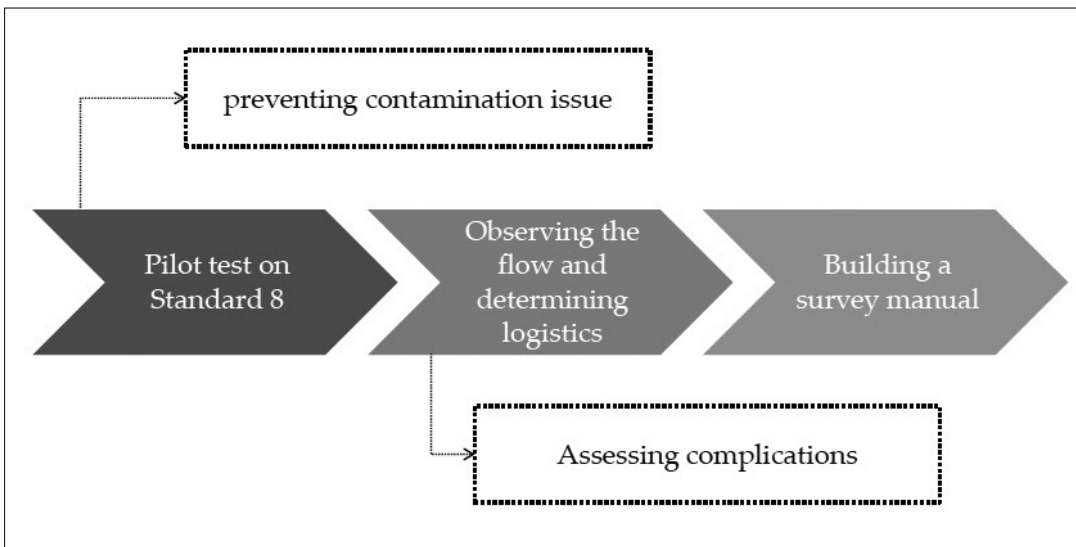
설문지는 여러 차례 시험조사(Pilot study) 통해 설문의 타당성을 검증하였다. 시험조사를 실제 조사대상 집단보다 나이가 어린 집단을 대상으로 실시하여 응답자들의 설문내용에 대한 이해도를 확인하고, 설문내용에 사전에 노출됨으로써 발생할 수 있는 문제(Contamination issues)를 방지하였다. 또한, 설문내용의 타당성뿐만 아니라, 설문을 수행하는 이동 경로, 수단 등 전반적인 현장조사의 흐름을 직접 시험해 봄으로써 실제상황에서 발생할 수 있는 혼란을 최소화 하였다. 이와 더불어 설문 수행의 일관성을 확고히 하기 위해 설문요원들이 따라야 할 현장조사 매뉴얼을 시각자료를 포함하여 이해하기 쉽게 제작하고, 사전교육을 실시하였다.

〈 그림 7 〉 현장조사 절차



출처 : 김현철 PM 발표자료

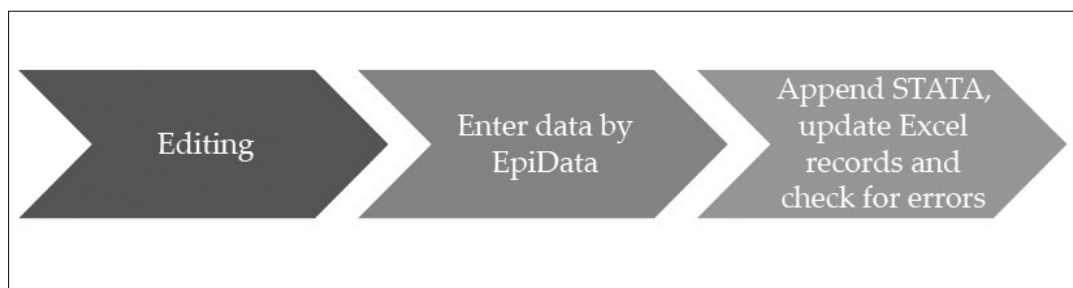
〈 그림 8 〉 시험조사 절차



출처 : 김현철 PM 발표자료

자료의 입력은 설문조사를 실시한 현장에서 동시에 하는 것이 가장 이상적이나, 현실 여건상 설문조사를 완료하고 수행하였다. 자료의 체계적인 정리를 위해 자료입력 프로그램(EpiData)을 활용하였다. 또한 STATA, Excel 등을 활용하여 오류가 있는지 여부를 확인하였다.

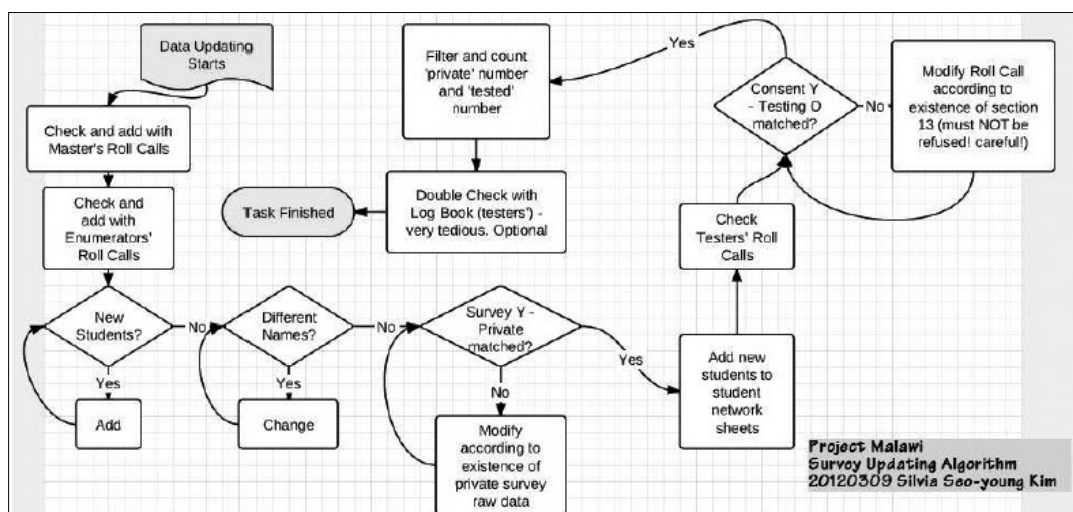
〈 그림 9 〉 데이터 입력 절차



출처 : 김현철 PM 발표자료

모니터링을 통해 동일한 성과지표에 대한 자료를 지속적으로 수집하였는데, 현장 조사시에 매번 학생들의 출석여부를 확인하여 새로운 학생의 유입 또는 유출을 확인하였다. 이는 사업대상지의 특성상 동일한 응답자더라도 이름이 바뀌는 경우가 있어 일반적인 정보의 업데이트가 필요하였다. 동일한 응답자에 대한 추가 데이터를 지속적으로 업데이트 하였으며, 자료의 업데이트 절차를 아래 그림과 같이 도식화 하여, 모든 데이터 관리자가 자료 정리를 위한 가이드라인으로 활용하도록 하였다.

〈 그림 10 〉 데이터 업데이트 절차



출처 : 김현철 PM 발표자료

데이터 정리(Data cleaning) 단계에서는 현지 조사요원들이 1차적으로 입력한 데이터를 자료입력 프로그램을 통해 분포도를 그려봄으로써 특이성을 보이는 데이터가 있는지 확인하고, 표준에서 크게 벗어나 있는 사례(Case)는 설문지를 다시 확인 하는 방법으로 데이터를 검증하였다. 또한 수집된 데이터의 정확성을 확고히 하기 위해 2차적으로는 수집된 데이터를 Excel 파일로 출력하여 직접 눈으로 확인 하였다.

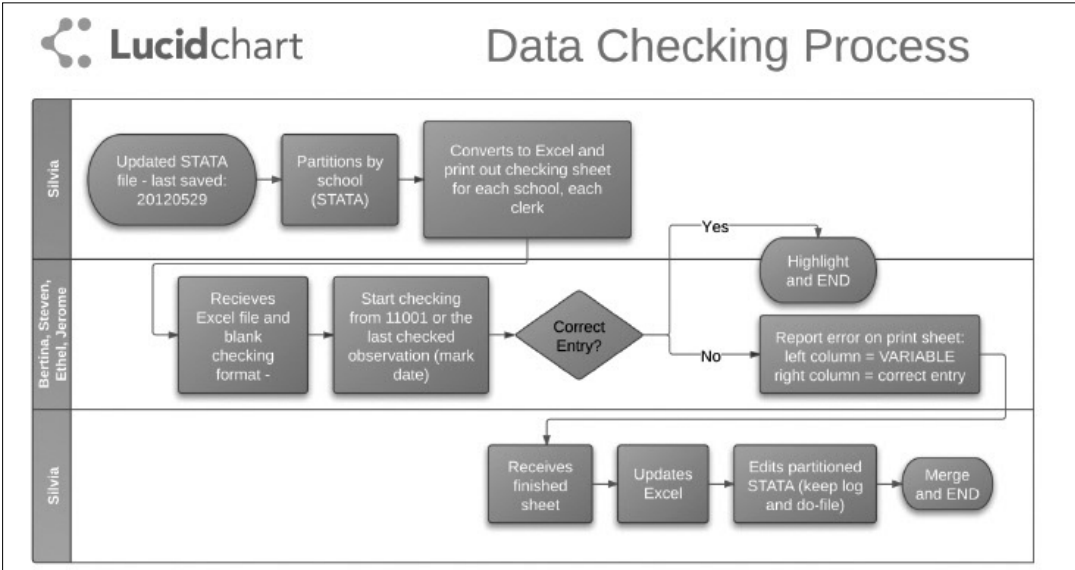
〈 그림 11 〉 현장조사 절차



출처 : 김현철 PM 발표자료

이와 같은 데이터 점검 절차 또한 개인차를 최소화하고 전반적인 과정을 표준화 하기위해 데이터 점검 절차를 아래 그림과 같이 도식화 하여, 가이드 라인으로 활용하였다.

〈 그림 12 〉 데이터 점검 절차



출처 : 김현철 PM 발표자료

이와 같은 방법은 일반적인 자료수집 방법에 비해 시간과 노력, 예산이 많이 요구되어 이와 같은 방법을 모든 사업에 엄격하게 적용하는 데는 한계가 있다. 그러나 이러한 자료조사 방법을 통해 생산된 신뢰할 만한 기초자료는 질 좋은 평가를 위한 기반을 제공한다. 또한 정확한 기초 자료를 활용하여 도출된 평가 결과는 성과의 계량적 측정을 가능케 하고, 향후 유사사업에 참고할 만한 유용한 자료로 활용할 수 있다. 따라서 이와 같은 현장조사의 모범사례를 시범적으로 적용해 볼 필요가 있다.

### 3. 향후과제 및 연구제언

앞에서 살펴 본 바와 같이 개발평가의 품질관리는 평가의 전 과정에서 이루어지는 것이 중요하다. 또한 평가의 품질 관리를 위한 세부사항 점검뿐만 아니라 체계적인 평가설계 및 정확하고 신뢰할 만한 데이터 수집은 평가의 품질보증에 기반이 된다. 즉, 체계적인 성과관리는 평가의 질 확보에 기여할 수 있으므로, 두 가지 개념은 유기적인 관계로 이해되어야 한다.

이와 같은 맥락에서 성과관리 필요성에 대한 강조와 더불어 지속적인 대내외 교육을 통해 품질관리의 필요성 및 내용을 확산할 필요가 있다. 가령, DAC의 품질기준에 대한 인식을 높여 이를 평가의 품질보증을 위한 일종의 체크리스트로 활용하도록 해야 한다. 현재 KOICA의 평가 품질관리 기준은 평가실에서 수행하는 평가에 한해서 엄격하게 적용되고 있다. 그러나 평가자 및 평가 관리자들의 품질보증 의무를 확대하기 위해 점진적으로 동 기준의 적용을 KOICA 내에서 실시하는 모든 평가로 확대할 예정이다.

영향평가와 같은 과학적인 평가 설계는 일반적인 평가에 비해 많은 인적·물적 자원과 시간의 투입이 요구되므로, 이와 같은 방식을 모든 사업에 포괄적으로 적용하기는 어렵다. 그러나 이러한 평가는 정확한 사업성과 측정을 가능케 할 뿐만 아니라 신뢰할 만한 평가기반 자료를 생성하여 결과적으로는 양질의 평가를 도출할 수 있게 한다. 평가의 주요 목적이 책임성 확보와 학습이라는 점을 감안했을 때 영향평가는 평가 본연의 목적을 달성함과 동시에 평가의 품질보증을 확보할 수 있는 적합한 방법으로 고려될 수 있다. 따라서 사업모듈 검증, 양질의 평가결과 도출 등이 필요할 경우에는 영향평가를 통한 성과입증을 시범적으로 시도할 수 있다.

최근 성과관리에 대한 중요성이 대두되면서 KOICA에서는 2010년부터 성과관리 시범사업을

선정하여 추진하고 있다. 대부분의 시범사업이 초기단계에 머무르고 있어, 현재까지는 사업성과 모형 수립, 성과지표 및 자료수집 구체화 등 성과계획 수립에 초점을 두고 추진되어 왔다. 성과관리 시범사업의 내실 있는 평가를 위해서는 사업수행기간 중에 성과지표와 관련한 데이터 수집을 지속적으로 시행해야 하며, 이러한 측면에서 아프리카 미래재단의 현장조사 사례는 좋은 참고사례로 활용될 만하다.

향후 우리는 위에서 언급된 지식을 활용한 평가 품질보증노력 이외에도 평가 모범사례 도출 및 노하우 확산을 통해 평가의 질 제고를 위한 지속적인 노력이 필요하다. 또한 평가 전 과정에 걸쳐 품질관리를 의무화하기 위한 제도적 장치가 개발이 되어야 할 것으로 생각된다.